8th International Command and Control Research and Technology Symposium

June 17 – 19, 2003

Title: **Using Gaming and Agent Technology to Explore C2**

Track: **C2 Experimentation**

Authors:

Peter P. Perla

Director for Interactive Research Products

Center for Naval Analyses

4825 Mark Center Drive

Alexandria, VA 22311

Tel: 703 824-2357

Fax: 703 824-2949

Email: perlap@cna.org

Julia Loughran

President

ThoughtLink, Inc.

2009 Cantata Court

Vienna, VA 22182-3313

Tel 703-281-5694

Fax 703-319-8196

loughran@thoughtlink.com

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **JUN 2003**	2. REPORT TYPE	3. DATES COVERED **00-00-2003 to 00-00-2003**

4. TITLE AND SUBTITLE **Using Gaming and Agent Technology to Explore C2**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Center for Naval Analyses,4825 Mark Center Drive,Alexandria,VA,22311**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S)
	11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**		**68**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

USING GAMING AND AGENT TECHNOLOGY TO EXPLORE C2

DR. PETER P. PERLA

Center for Naval Analyses
4825 Mark Center Drive
Alexandria, VA 22311
703-824-2357
perlap@cna.org

MS. JULIA J. LOUGHRAN

ThoughtLink, Inc.
2009 Cantata Court
Vienna, VA 22182
703-281-5694
loughran@thoughtlink.com

ABSTRACT

Traditional command and control research occurs at two extremes of the cost and fidelity spectrums. At one end, low cost seminar games and simple abstractions, like chess, offer insights, but lack rigorous scientific techniques for analysis. On the other hand, highly detailed simulations, like those conducted for the US Navy's Global War Game, cost time and money, and offer little in support of developing scientific proofs. This paper details the methodology of employing two complementary concepts to the field of C2 research: *game-based experimentation* using distillation games, and *agent-based* methods. These approaches fall midway on the cost and fidelity spectrums. The game distillation, *SCUDHunt*, has proven to be successful in providing a rigorous scientific and statistical approach for experimentation. The results of *SCUDHunt* experiments offer insights into team behavior, shared situational awareness, and team performance. To complement this human-player environment, we created *SCUDHunt* computer agents. This agent-based approach provides an exploratory environment complementary to the human-based game. This paper provides an overview of the work we, and others, have done in these areas to date, and proposes some future directions to develop the promise of this approach.

EXECUTIVE SUMMARY

Games provide a wealth of flexibility for exploring, testing, and demonstrating a host of variables and issues associated with command and control (C2). Unfortunately, even a single iteration of a complex, multiplayer, large-scale operational wargame is expensive in time and money. Conducting multiple iterations of such wargames is impractical. Whatever their value may be for other purposes, such games are relatively poor vehicles for some forms of scientific experimentation—in particular, for hypothesis testing and developing "scientific proof."

There are, however, two complementary concepts that we have applied to some initial research, and that we believe to have great potential value for the future. The first of these concepts is game-based experimentation using distillation-style games. The

second is the use of agent-based methods to explore complex systems. Integrating these two techniques promises to be a powerful new approach to improving our understanding and analysis of command and control.

In the course of several research projects conducted over the last three years, we have developed a simplified, tightly focused experimental gaming environment called *SCUDHunt*. The *SCUDHunt* environment allowed us to tailor the design and mode of game play to focus on specific topics related to the shared situational awareness and performance of teams of human players. This approach allowed us—and other researchers—to formulate and test hypotheses using rigorous scientific and statistical techniques for experimental design and analysis. We characterize these sorts of games as *distillations*—distinguishing them from simple abstractions, like chess, and detailed simulations, like the U.S. Navy's Global War Game. Basing experimentation on distillation games allows researchers to conduct experimental design, data collection, and statistical analysis in ways not available for large exercises or demonstrations.

In the past year, we took the *SCUDHunt* experimental environment beyond the realm of human players playing the game. We created computer agents to play the game in a manner analogous to that of human players. We developed this agent-based approach from concepts underlying the "new sciences" of complex systems and cellular automata, sciences that explore whether the behavior of different complex systems may stem from some relatively small set of fundamental principles.

Agent-based exploratory models are based on the idea that complex global behavior can derive from simpler low-level interactions among components. The goals of building and using agent models include learning quantitative and qualitative properties of the real system and testing hypotheses about the origin of observed emergent properties. The fundamental technique of the approach calls for experimenting with initial conditions at the micro-level to generate desired behaviors at macro-level.

We conducted an initial mini-experimental campaign, integrating a human-based experiment, with an agent-based experiment. These experiments measured variables we associate with shared situation awareness (SSA) and accuracy of assessment. The agent-based model may help us better reflect the complexities of differences in human belief systems and trust for each other's judgments, but at this early stage of development, we have not been able to vary parameter values over a sufficiently extensive space to explore those dynamics in much detail. However, in both the human-based and agent-based experiments, information quality had an important effect on the accuracy of decisions.

Our application of agent-based techniques in the "*SCUDHunt* universe" allows us to leverage the power of agent technology to broaden and deepen our exploration of human behavior in our experimental environment. It is relatively easy to create agents and use them to play many iterations of the *SCUDHunt* game—far easier than recruiting and managing the same number of human agents for the same number of iterations. By using the results of one type of experiment as a "question generator" for the other, we can maximize the value of both. For instance, should an interesting situation arise in the human-based game, a similar situation can be created and explored in depth in the agent-based game. Likewise, if the behavior of the computer agents produces particularly intriguing results, we can explore the situation further using human players, to try to

understand whether and how the agent-based play reflects actual human activities. This mutual-feedback mechanism allows for the examination of a large variety of notional command and control architectures at minimal cost.

But it is from the combination of these two approaches that we feel we can get the highest payoff. Human game-based experimentation, through the implementation of distillations, is a scientifically based and statistically valid technique that can help us explore practical questions about human performance in C2-related tasks. Such insights are of fundamental importance if we are to improve our understanding and representations of such operational concepts as network-centric warfare, information warfare, and self-synchronizing command systems. The use of adaptive agent simulations within the context of game-based experimentation can help address one of the main difficulties of experimentation with human players: finding appropriate numbers and types of human players for the game. Using agent-based gaming will allow us to explore the experimental design space more thoroughly and much more quickly than is possible using games with live participants.

As we look to the future, we are struck by today's current rage for "transformation." DoD has established an office whose primary purpose is to advocate and pursue the transformation of the U.S. military establishment. Panels and study groups are convened and meet to report on whether new ideas are, or are not, transformational enough to be considered for future funding. To transform the way we act, however, we must first transform the way we think. Our work on this research has convinced us that, at the very least, we must transform our thinking about how to study and evaluate military command and control by integrating game-based experimentation and agent-based methods.

1. CHALLENGES IN C2 RESEARCH

The information revolution has affected C2 processes, systems, and the organizations that implement them. While these changes have increased the importance of C2 analysis, they have also increased the analytical challenges. Today, information technology is being used as a weapon and its effective employment can be a force multiplier[1]. Therefore, it is extremely beneficial to find ways to enhance our analytical approaches to C2.

In a recent issue of *Phalanx*, *the Bulletin of Military Operations Research*, Mr. Vincent P. Roske, Jr., Deputy Director, J8, wrote that the difficulty of command and control research "comes when trying to account for the creativity, initiative, and perception of the human factors."[2] Human beings produce emergent and adaptive behaviors and we need complementary approaches for analyzing these factors.

---

[1] http://www.dodccrp.org/2000CCRTS/ppt/10

[2] "Opening Up Military Analysis: Exploring Beyond the Boundaries," Vincent P. Roske, Jr., Deputy Director, J8 (Wargaming, Simulation and Analysis), The Joint Staff, *Phalanx: The Bulletin of Military Operations Research*, June 2002, p. 1

According to Roske, the most popular approach to open systems analysis has traditionally been wargaming. Unfortunately, in most instances these wargames are large, multi-day exercises, they take months to plan, run in real-time, try to capture too many experimental variables and are slow, expensive, and inefficient for gathering scientifically statistical results.

Traditional analytical methods have difficulty representing emergent and adaptive behavior. Many models and constructive simulations are designed for closed systems—meaning systems in which the variables can be controlled (e.g., weapon capabilities). These tools are also not well suited to C2 analysis because they do not represent human behavior and they cannot depict the complexities associated with network-centric and asymmetric environments.

Today, underlying information systems and human decision making play a greater role than sheer weapon power in winning the fight or gaining an upper hand on the enemy. This new information and decision rich environment requires an "open systems" approach to analysis. Old techniques usually involved controlling a system's variables; today's problems demand new techniques that allow us to study emergent behavior.

This paper discusses two new approaches for open system analysis. The first is the use of game-based experimentation using "distillation games"—games that reduce real-world problems and entities to simplified representations focused on a few prominent elements of the real-world environment.[3] The second is the use of agent-based methods. The authors and their teams, from the Center for Naval Analyses (CNA) and ThoughtLink, Inc., have successfully used these approaches to conduct C2 analysis. Our analyses have focused on team behavior and the factors affecting a team's ability to build shared situational awareness and to make quality decisions. Variations of *SCUDHunt*, a C2 distillation game, were used in both of these approaches.

The paper also discusses the benefits of using the combination of these two approaches to explore C2. We outline an approach for using this methodology in a C2 *experimentation campaign plan*, which is defined as an "organized way of testing innovations that allow refinement and support increased understanding over time."[4]

The benefits of using distillation games for C2 research and analysis are many. Some of these benefits include the fact that they provide powerful abstractions, they reduce the complexity of a high-fidelity real-world environment, they support statistical analysis, and they are fun—a characteristic that helps keep human participants engaged.

The benefits of using an agent-based approach include the facts that complex behavior can emerge from simple rules, they are easy to manipulate and can therefore cover a large section of the analytical landscape, and they eliminate the logistical headaches associated with conducting human-based experiments.

---

[3] CNA Research Memorandum (CRM) D0006277.A1, *Game-Based Experimentation for Research in Command and Control and Shared Situational Awareness*, by Peter P. Perla, Michael Markowitz, and Christopher Weuve, May 2002. Hereafter cited as Perla, 2002.

[4] *Code of Best Practice Experimentation*, David S. Alberts, Richard E. Hayes, DoD Command and Control Research Program, July 2002, p. 25

The remainder of this paper discusses our research, some implications, and some future directions.

Section 2 defines *SCUDHunt,* the distillation game co-developed by CNA and ThoughtLink for our C2 analysis.

Section 3 provides a summary of the human-based *SCUDHunt* experiments conducted to date.

Section 4 discusses the application of an agent-based approach to *SCUDHunt* as a proof of principle.

Section 5 presents the analysis of a mini-experimental campaign that integrated human- and agent-based experiments using *SCUDHunt*.

Section 6 proposes a way ahead for using these approaches, either individually or combined, in support of a robust C2 experimental campaign plan.

2. SCUDHUNT: THE GAME

SCUDHunt is a simple distillation game of command and control, played over the Internet by (generally) distributed teams. The game was co-developed by CNA and ThoughtLink, Inc. *SCUDHunt* is similar to the popular game *Battleship*, in which a player hides a fleet of warships on a grid while his opponent explores sections of that grid in an attempt to sink those ships. In *SCUDHunt,* however, players play cooperatively on a single team trying to determine (within a specified number of turns) where three SCUD launchers are hidden on a 5 X 5 grid. Launchers are randomly hidden on the map grid at the start of each game and these launchers remain stationary throughout game play.

The game's operational back-story states that players (generally four-person teams) are part of a joint or combined force and their team's objective is to locate the three stationary SCUD launchers hidden in the hostile country of Korona. Players command one or more information, surveillance, and reconnaissance (ISR) assets, with different capabilities and different SCUD-detection probabilities. These probabilities are described in a general way to players in on-line asset briefings they receive before the game is launched. During the game, players must collaborate with each other and share information in order to build a shared picture of where the SCUD launchers may be hidden. The mode of communication, the type of visualization, and the asset detection probabilities may vary depending on the experimental conditions.

Player positions and assets are:

- Space Asset Manager: controls the reconnaissance satellite;
- Intelligence Manager: controls the communications intelligence (COMINT) and human intelligence, the spy (HUMINT);
- Air Asset Manager: controls the manned aircraft and the unmanned air vehicle (UAV), and

- SpecOps Manager: controls the special operations forces (Navy Seals and the Joint SpecOps team).

The game—whose success often depends on the team's developing an accurate shared picture of the information contained on the game board—was originally designed as an experimental test bed for research into shared situational awareness (SSA). The sponsor for the initial research was the Defense Advanced Research Projects Agency under the program entitled Wargaming the Asymmetric Environment (WAE). The definition of SSA we used in this work was proposed by Mica Endsley in 1995: "the perception of the elements in the environment within a volume of space and time, the comprehension of their meaning, the projection of their status into the near future, and the prediction of how various actions will affect the fulfillment of one's goals."[5]

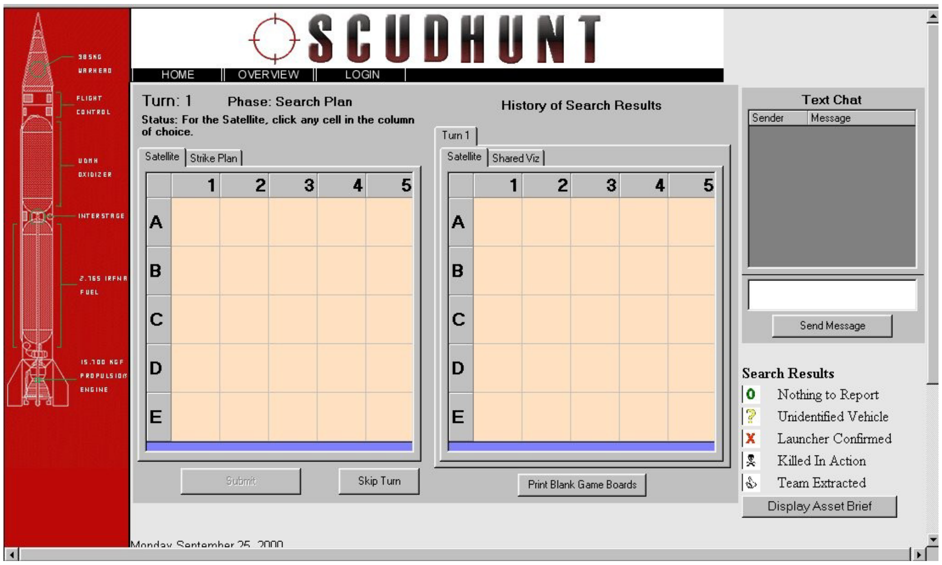
Although each team member believed the objective of the *game* was to locate the hidden SCUDs, the undisclosed objective of the *experiment* was to gather information that would provide insights into the player's situation awareness, in which the "situation" is the location of the SCUDs, and the "situation awareness" constitutes the individual's belief (or guess) as to the locations of the three SCUDs. The measurement of "shared situation awareness," in turn, reflects the collective overlap in each of the individual team member's awareness at various points throughout the game. The quality of their decisions (or accuracy) is determined by the team's ability to identify all of the hidden SCUDs. This measure can also be applied to characterize the accuracy of individual players.

The *SCUDHunt* game board is shown in figure 1 below. The left-hand grid square is used to place assets and submit a strike plan. The right-hand grid square presents the results from the assets search (and may include the results from other assets under the Shared Viz (visualization) option) and the farthest right-hand window shows the text chat window, which is one of the communication conditions that can be used in the game.

---

[5] Endsley, Mica, "Towards a Theory of Situation Awareness in Dynamic Systems," *Human Factors* (1), 1995, pp. 32-64

Figure 1: *SCUDHunt* game board



During game play, players gather information by positioning their search assets on the 5X5 grid game board. Some of the assets are limited to searching one grid square at a time (e.g., the Navy Seals) while others, like the reconnaissance satellite, can search multiple grid squares in a single turn. After all team members have placed their assets, each individual asset's findings are returned to the appropriate asset manager. Players then have to share their search results to form a complete picture of the overall results for a given turn.

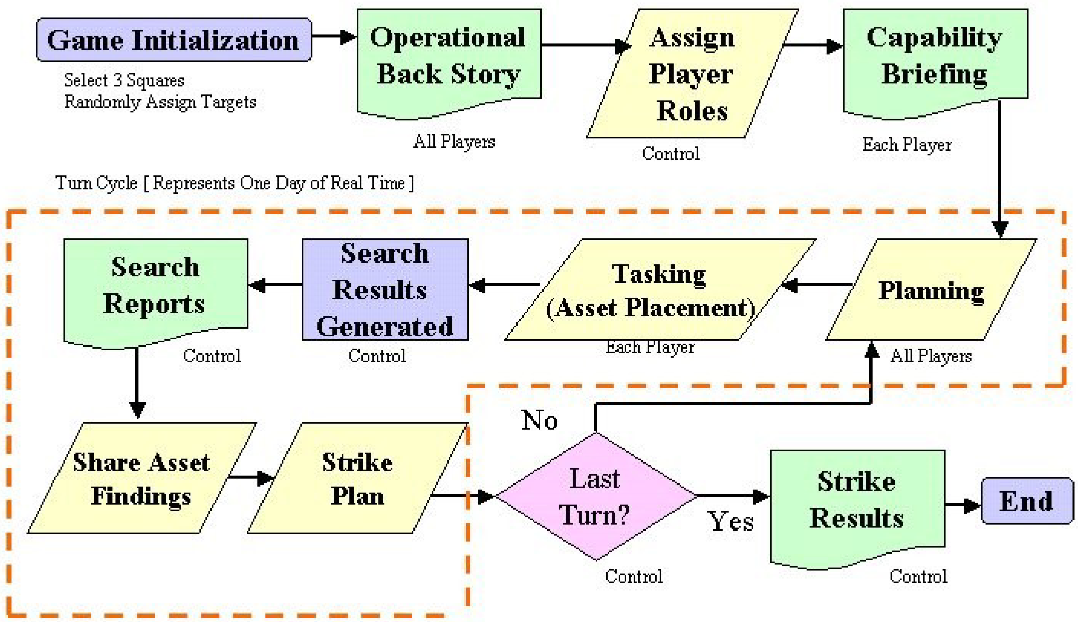
Three basic search results are returned: 0, when there is nothing significant to report; ? when vehicles are detected but cannot be confirmed as SCUD launchers, and X, when a launcher is detected. Some sensors can be killed or temporarily disabled. Search results may be accurate or erroneous based on the detection probabilities of each of asset and the random number drawn on each turn for that asset. Depending on the reliability of an asset, a result '0' or 'X' may not be correct. An incorrect '0' is a *false negative* (meaning there was actually a launcher in that square) and an incorrect 'X' is a *false positive* (meaning there was actually no launcher in that square). Other results may indicate that an asset was either killed or shot down, or a that a team was extracted. The frequency of false positives and false negatives is a factor we have investigated in terms of its relationship to team SSA and accuracy.

Most games incorporate some form of communication among the players, allowing them to share the results of their searches with each other. (Although we have conducted games in which such communication was prohibited.) Communication conditions in the various *SCUDHunt* experiments included Internet-enabled text chat, group teleconferences, or the use of shared visualization tools.

After team members compile their own mental model (picture) of the situation, they are asked individually to report their best guess of where the SCUD launchers are located, nominating a minimum of three grid squares. While no upper limit was set as to the number of squares specified, players were told to identify the fewest number of

squares that would still represent their beliefs about the locations of the SCUDs. We use these individual strike recommendations to compute a shared situational awareness score for the team. Each turn ends with all players voting for three or more grid squares. The typical game lasted for five turns (although this varied on a per experiment basis). The overall flow of the game is shown in figure 2 below.

Figure 2: *SCUDHunt* Game Flow



The game is instrumented so that each player's actions, the experimental settings, as well as the communication (if text chat, shared visualization, or push visualization is used) are captured in a Microsoft Access database.

The measure we used to quantify SSA is the overlap in launcher location assessments (strike recommendations) among team members, regardless of whether their assessment is right or wrong. The team's SSA score is calculated as the ratio of the total number of target squares recommended by all players to total number of unique squares designated. If a team has perfect SSA, for example, all four team members vote for the same three squares, which gives a score of 12 (total number of votes) divided by 3 (the total number of unique squares), for a perfect SSA score of 4. An example of the lowest possible SSA score would be if all four team members vote for three different squares which would produce a score of 12 (total number of votes) divided by 12 (total number of unique squares) for an SSA score of 1.

We have explored several measures of accuracy in our experiments, but the most easily understood of those measures is simply the fraction of recommended squares that actually contained SCUD launchers. The accuracy score for a player or a team is calculated as the ratio of recommended squares that actually contained SCUD launchers to the total number squares nominated. An example of perfect team accuracy would be if

all four players vote for the same three squares, each of which actually contained a launcher. In this case, the accuracy score would be 1.0. The lowest possible accuracy is 0, which occurs if the team does not identify any launcher squares.

3. SCUDHUNT EXPERIMENTS AND FINDINGS

We developed *SCUDHunt* in 2000 to support a DARPA program by studying factors influencing a team's shared situational awareness. The original experiment used a Latin square design to explore how different modes of communication and visualization affect a distributed team's SSA. In addition to producing data amenable to statistical analysis, and some interesting statistically significant results, the experiment was also a success because it saved considerable amounts of time and money when compared to traditional analytical approaches. Because the game was implemented in Visual Basic, an easy-to-use programming language with shareware tools and other low-cost web technologies, the experiment cost only thousands—rather than millions—of dollars. Table 1 provides an overview of the various *SCUDHunt* experiments that CNA, ThoughtLink, and other organizations—particularly the Naval War College—have conducted to explore concepts of information superiority, training, and leadership.

Table 1: Summary of *SCUDHunt* Experiments[6]

Experiment/Year	Conducted by	For	Experimental Variables
Experiment #1; 2000	ThoughtLink and CNA	DARPA	Availability of visualization, type of communication
Data Mining of Experiment #1; 2001	ThoughtLink	Joint C4ISR Decision Support Center	Data mining of original experiment for quality of decisions
Experiment #2; 2002	George Mason University	Army Research Institute	Training on own or all assets, mode of communication
Experiment #3; 2002	Naval War College, CNA, ThoughtLink	Naval War College	Command method, type of visualization
Experiment #4; 2002	ThoughtLink, Naval War College, CNA	Joint C4ISR Decision Support Center	Quality of information, type of visualization
Experiment Meta-Analysis; 2002	ThoughtLink	Joint C4ISR Decision Support Center	Meta Analysis of four *SCUDHunt* experiments

[6] In addition, the University of Arizona conducted a study using *SCUDHunt* in 2001 looking at leadership and knowledge of sensor reliability. This study was not listed because the authors do not have any additional information regarding this experiment.

The experimental variables that have been of interest in these *SCUDHunt* experiments include:

Availability of visualization: This variable included whether participants saw a shared visualization screen with all of the aggregated results or only saw the results from their own assets.

Type or Mode of communication: This variable looked at differences in communicating via a team teleconference or via a shared text chat window.

Training on own or all assets: This variable included whether or not team members were trained in the capabilities of all the information assets or just their own. Knowledge (all vs. own) was manipulated between teams and concerned the training content provided to the players regarding characteristics (mobility, reliability, vulnerability, etc.) of the assets used to collect intelligence regarding SCUD missile launcher positions. Players received limited preliminary training on all assets. In the all-knowledge condition, preliminary training touched on all assets briefly, then individual players received training focused on the assets they would control during the game, followed by training focused on the assets controlled by the other player. In the own-knowledge condition, players only received training focused on assets they would control during the game. Manipulation of own- and all-knowledge were intended to affect the content of the shared mental models that the players had at the beginning of the game.

Command methods: this variable explored three styles of command method: command by direction, command by influence, and command by plan[7]. In the command by direction condition a fifth player, a commander, gave specific orders to each of the four sensor players for where to place their assets each turn. In the command by plan condition, an overall plan was promulgated by the control group acting as a higher command authority, with branches and options for how the players were to proceed with their search. In the command by influence condition, an overall mission was defined (in simplest terms, to find the SCUD launchers) and the players were left free to coordinate among themselves about how best to carry out their mission.

Type of visualization (in Experiments #3 and #4): This variable refers to the use of either shared visualization or post visualization, a concept introduced to *SCUDHunt* by the Naval War College. In the shared visualization condition, all sensor returns were given to all players. In the post visualization condition, players were asked to post to a shared display their *interpretations* of the sensor returns.

Quality of information: Quality of information translates to the reliability with which each asset can identify a hidden SCUD launcher.

- Medium QOI. Probabilities are the same as in all prior *SCUDHunt* experiments. This represents the base case. There is a small chance of false positives (an X returned in an empty square) and false negatives (a 0 returned in a launcher square).

---

[7] Command and Control at the Crossroads. *Parameters*, Autumn, Czerwinski, T.J., 1996, pp.121-132, or on-line at http://carlisle-www.army.mil/usawc/Parameters/96autumn/czerwins.htm

- High QOI. False negatives are decreased. False positives are unchanged from the base case.
- Low QOI. False positives are increased. False negatives are unchanged from the base case.

SCUDHunt has proven to be a flexible experimental testbed. It has been used for a variety of experimental conditions, ranging from factors influencing distributed training to visualization and communication modes to command methods. Team size has varied, from two-person to four-person teams. Other factors that can easily be varied include the use of text chat, the number of turns in the game, and the detection probabilities used for each asset.

Key findings from these experiments include:

- Mode of communication is not as important to SSA and quality as the fact that there is communication.
- There is no big difference between shared (raw) vs. post (interpreted) visualization.
- There is a fairly strong relationship to team SSA and accuracy.
- Good quality of information leads to high SSA; even moderate degradation of info quality degrades SSA
- Teams matter; we want to further explore elements of team composition and team dynamics.
- We see mixed statistical results about a learning effect based on the number of games played, but the players themselves have a strong perception of a learning effect.

Below are brief descriptions of each of the experiments and their major findings. These summaries are not intended to provide the specific details of the experiments, but instead, illustrate the flexibility that a distillation game can bring to exploring a complex issue like C2.

3.1 EXPERIMENT #1: THE ORIGINAL EXPERIMENT

ThoughtLink and the Center for Naval Analyses conducted the original experiment in 2000 for DARPA. We assessed how different modes of communication (three levels: none, text chat, audio) and visualization (two levels: none, shared vis) affected a distributed team's ability to develop and maintain shared situational awareness.

Six four-person teams each played six online games of *SCUDHunt* in different order based on a Latin Square experimental design. Players filled out pre-game questionnaires concerning background, and post-game questionnaires about their experiences during each game played. Results indicated that communications and shared

visualization affected SSA, but mode of communications did not seem to matter (so long as there was one: text chat or phone).[8]

3.2 DATA MINING OF EXPERIMENT #1

The original SSA experiment produced a wealth of data that subsequently was mined to identify other factors affecting team decision-making. For instance, does the mode of communication or use of a shared-visualization tool affect the quality of decisions? Quality of decisions, or accuracy, was determined by the fraction correct, as described earlier.

The data mining used regression analysis and standard analysis of variance techniques to explore possible relationships in the *SCUDHunt* data between:

- Team quality and communication and visualization modes
- Team quality and SSA
- Individual quality and asset type
- Individual quality and player's subjective assessments of the games
- Order of games

The findings from this analysis showed that the availability of any form of communications, either direct (through text chat or voice) or indirect (through shared visualization) was the key difference affecting the quality of decisions. The only characteristic that appears to affect an individual's quality score is team capability (i.e., individuals do well on teams that do well). The data mining raised a number of issues to be explored in future research, including an analysis of the interplay between communication mode and shared visualization, what factors influence "good teams," and what other player or leadership characteristics can be found in both individuals and teams that make high quality decisions.[9]

3.3 EXPERIMENT #2: ARI/GMU EXPERIMENT

In early 2002, George Mason University (GMU) conducted a study in conjunction with, and for, the Army Research Institute (ARI). The purpose of this experiment was to determine whether cross-training (training someone on related tasks as well as their own tasks) improves team performance. The two experimental conditions were: knowledge of assets (two levels: own assets, all assets) and mode of communication (two levels: voice, chat). Two-person teams worked to locate missile launchers.

Accuracy and SSA scores were the primary performance measures. Other variables of interest were measures of communication (number and type of messages)

---

[8] Detailed results of Experiment #1 are reported in CNA Research Memorandum (CRM) D0002722.A1, *Gaming and Shared Situation Awareness*, by Peter P. Perla, et al., November 2000.

[9] More detailed information on the data-mining project can be found in *Key Drivers for C2 Performance: Data Mining SCUDHunt Experiment Data*, by Julia Loughran, Marcy Stahl, and Peter P. Perla ThoughtLink, Inc. report for Joint C4ISR Decision Support Center, November, 2001.

during each of five turns and during each of two games, measures of perceived effort, and post-game measures of knowledge. Questionnaires were given before and after games concerning demographics (gender, age, and computer skills), motivation, social collaboration, and post-game reactions to the game and training.

At the time of this writing, a final report concerning results of Experiment 2 has not yet been completed. In addition, George Mason University and the Army Research Institute are currently conducting another experiment using *SCUDHunt.*

3.4 EXPERIMENT #3: THE NAVAL WAR COLLEGE'S EXPERIMENT

Experiment #3 was conducted by the Naval War College (with analytical support from CNA, and technical support from ThoughtLink) in the spring of 2002. It concerned effects of command method (three levels: command by plan, by influence, by direction) and visualization type (two levels: shared vis, post vis) on team SSA and accuracy scores. Three styles of command method were investigated: command by direction, command by influence, and command by plan. Six four-person teams were tested under each of the six experimental conditions using a Latin Square experimental design.

Results indicated a statistically significant improvement in both SSA and accuracy scores of teams employing a command-by-direction style when compared to the same teams playing under command-by-influence or command-by-plan styles.[10]

3.5 EXPERIMENT #4: QUALITY OF INFO AND POST VIS

ThoughtLink, with extensive support from the Naval War College and CNA, conducted this experiment in the summer of 2002 for the Joint C4ISR Center. The experiment concerned the effects of quality of information (three levels: high, medium, low) and visualization type (two levels: shared vis, post vis) on team SSA and accuracy scores.

Six four-person teams were tested under each of the six experimental conditions using a Latin Square experimental design. The strongest data from this experiment was that teams matter and team differences have a strong effect on accuracy and SSA scores. Another interesting finding was that although quality of information strongly affects accuracy, it has little effect on SSA. We will discuss this experiment further in section 5, particularly in relation to the agent-based approach.[11]

---

[10] For more detailed reports on Experiment 3, see CNA Research Memorandum (CRM) D0006277.A1, *Game-Based Experimentation for Research in Command and Control and Shared Situational Awareness*, by Peter P. Perla et al., May 2002. Hereafter cited as Perla, 2002.

[11] For details about this experiment see *Exploring Joint Force Command and Control Concepts Using SCUDHunt – Final Report*, by Marcy Stahl and Julia J. Loughran, ThoughtLink, Inc. report for the Joint C4ISR Decision Support Center, October 2002. Hereafter cited as TLI, 2002

3.6 META ANALYSIS OF SCUDHUNT EXPERIMENTS

Since the first four *SCUDHunt* experiments involved some common independent variables, and the dependent variables (performance measures) are comparable, it was considered useful to conduct a meta-analysis in order to examine some specific relationships. Of special interest is the relationship between SSA and accuracy scores.

The purpose of this study was to conduct a meta-analysis of *SCUDHunt* experimental data to assess important relationships between team and individual characteristics and game performance measures, derived from suggestions made by previous research. Of special interest in this study are relationships between:

- Team SSA and team accuracy scores,
- Subjective measures of SSA and team accuracy,
- False positives and false negative sensor asset reports, accuracy, and SSA scores
- Individual player characteristics and accuracy scores.

In addition to generating the statistical analysis for SSA scores, the game environment encouraged subjective observation of the activities of distributed teams, including:

- Playing the game appeared to promote bonding and trust among team members who had never met previously;
- Some female players appeared to have a higher degree of concern over reaching a team consensus; and
- Teams that developed repeatable (shared) processes of play appeared to have better shared awareness.

4. APPLICATION OF AGENT-BASED GAMES

Some of the problems plaguing all experimentation focused on human behavior include the need for a pool of test subjects and the development of appropriate protocols—and possibly even formal review boards—to ensure that the subjects have given properly informed consent to participating in the experiments. The experimental design we used for our previous *SCUDHunt*-based research involved 24 human players, in 6 teams of 4 players each, whose schedules had to be coordinated to accomplish the required sequence of game events. A potentially useful alternative to human-only experimentation is to integrate artificial game-playing agents into the mix. Agent-based games can play at least two major roles in such an integrated program of research.

- We can conduct exploratory research to identify potentially interesting patterns of behavior, which we could then probe more deeply with targeted human experimentation.
- We can observe how human players play the game and use agents to explore some of the possible underlying causes faster and more thoroughly.

Our application of agent-based concepts to game-based experimentation in general and to *SCUDHunt*-based research in particular grew out of the emerging sciences collectively known as complex systems. In particular, CNA's previous experience with cellular automata and agent-based models led us to a particular approach.

The emerging new sciences, often referred to as the study of complex systems, focus on exploring to what extent the behavior of different complex systems may depend on a set of fundamental principles. By understanding those fundamental principles, scientists hope to unlock the key to understanding the overall behavior of complex systems in ways not available to traditional approaches.

One of the simplest mathematical representations of a broad class of complex systems is the concept known as *cellular automata* (CA). CA systems have demonstrated their potential as powerful conceptual engines to study pattern formation in chemical reaction-diffusion systems, crystal growth, and the flow of vehicular traffic. They have proven useful idealizations of the behavior of physical fluids, neural networks, natural ecologies, and military C2 systems. This latter application first attracted our attention in the context of *SCUDHunt*. Our game-playing agents are not exactly cellular automata, but much of their creation derives from similar ways of thinking about modeling human behavior using simple, yet powerful, reductions of complex processes into simple decisions.[12]

4.1 AGENT-BASED SCUDHunt

Drawing on the philosophy underlying CA, our view of agent-based models is based on the notion that complex global behavior may derive from simpler, lower-level interactions among the components of the system. "Insights about the real-world system that the agent-based simulation is designed to model can then be gained by looking at the emergent structures induced by the interaction processes taking place within the simulation."[13]

Rather than building an agent-based simulation of the "real world," we built an agent-based simulation of the *SCUDHunt* universe, a distillation of a real-world environment focused on issues related to shared situational awareness and cooperative decision making.[14]

In addition to demonstrating the practicality of building such a set of game-playing agents, our goals in this process are well described by the motivations behind the use of agent-based models of the real world.

> The purpose behind building an agent-based simulation of [a] real-world system is twofold: it is to learn both the quantitative and qualitative properties of the real system. Agent-based simulations are well suited for

---

12 For a more detailed discussion of these ideas, see Andrew Ilachinski. *Cellular Automata: A Discrete Universe*. New Jersey: World Scientific, 2001. This section is derived mainly from chapter one, pp. 1–20.

13 Ilachinski, p. 564.

14 For a discussion of games in terms of abstractions, distillations, and simulations, as well as the general concept of game-based experimentation, see Perla, 2002.

> testing hypotheses about the origin of observed emergent properties in a system. This is done simply by experimenting with sets of initial conditions at the micro-level necessary to yield as set of desired behaviors at the macro-level.[15]

The first step in applying these techniques to the sorts of issues we originally created *SCUDHunt* to explore was to develop an approach to modeling—to a first order of representation—the decision-making behavior of human players of the game.

The basic idea behind the agent-based *SCUDHunt* system is to develop game-playing agents (a set of software routines) to represent the players of the standard *SCUDHunt* game. These agents should do the same things that human players do when they play *SCUDHunt*—collect and interpret information from and about the sensors they control, make decisions about where to place their sensors, and exchange that information and those decisions with each other. They also should make individual decisions about which grid squares they would "recommend" as the most likely target locations at the end of each turn of the game.

At the highest level of player interaction, a schematic of the overall *SCUDHunt* agent model looks like figure 3. Each individual agent stores and processes information. This information takes the form of their understanding about the capabilities and deployment restrictions of the sensors, their beliefs about the locations of actual SCUDs, any constraints they might have about communicating with each other, and the meaning (or possible range of meanings) of each search result from each sensor.

Based on their information and their assessment of it, the agents carry out the game actions. First, they process the information available to them to "update" their beliefs about the locations of the targets. Based on those beliefs, they have to make their "strike recommendations." Finally, in cooperation with the other agents, they decide where to place their search assets for the coming turn. Schematically, the job of the individual agents looks like figure 4.

---

[15] Ilachinski, p. 564.

Figure 3: Schematic of agent interaction

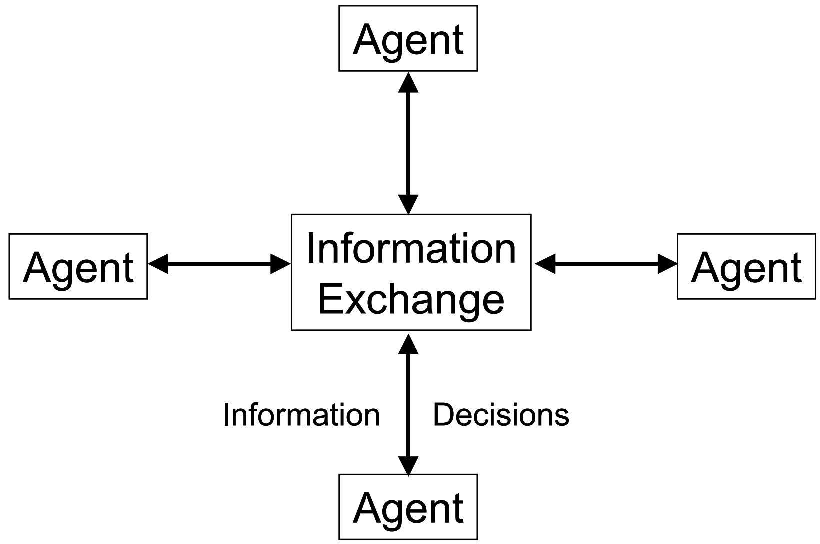
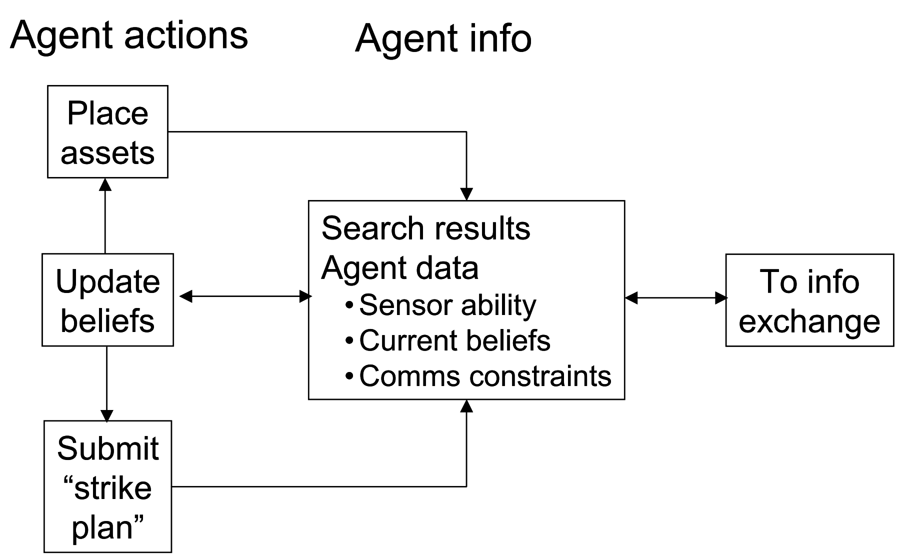


Figure 4: Schematic of individual player actions



The challenge is to endow agents with a personality-driven artificial intelligence that is simultaneously powerful enough to mimic some important aspects of human decision making (so that the agents' actions appear to be intelligent actions), and simple enough so that the analyst is not overwhelmed by having too many parametric "knobs" to tweak.

The elements of this mathematical model are described in detail in CNA's earlier report on this research, from which much of this and the following section is extracted.[16] The key components of the model include representations of each agent's:

- *Belief Matrix*, which represents the strength of the agent's belief that a target is, or is not, present in a specific grid square
- Interpretation of sensor reports and how they change his belief value for the grid squares
- Trust (of other agents) and how that affects the way he integrates the information they provide into his own belief calculations
- Strike-plan logic, the determination of which targets to recommend for strike
- Sensor-placement logic, the process of deciding where to place the agent's sensors to maximize some "fitness function" representing the various, possibly competing, motivations an agent may have as he decides how to allocate his search effort.

The belief matrix is the critical component of the *SCUDHunt* agent design—all decisions regarding sensor placement and strike plans are functions of it. The way in which the belief matrix changes, for a given agent A, is unique to A, and is a function of A's *personality*.

An agent's personality consists of the parameters that define how an agent obtains, interprets and uses game-generated information. These parameters are grouped according to the list above.

The first part of an agent's personality consists of parameters that define how an agent interprets reports from his own sensors (embodied in the *Sensor-Report: Launcher-Correlation Matrix*). The second component of A's belief matrix is the set of partial-beliefs stemming from reports communicated to A by agents to whom he is linked. This calculation involves an *Agent-to-Agent Trust Matrix* to account for the extent to which agent "i" trusts information communicated to him by agent "j".

A's belief matrix is updated according to the kind of information that is communicated to A by agents linked to him. Linked agents may act as simple conduits of raw information, and provide A with their own (unfiltered and uninterpreted) sensor reports. Alternatively, linked agents may provide A with their own *interpretations* of what their sensors reported to them (i.e., they pass to A their partial beliefs). A updates his own partial-belief according to these interpretations, not the raw data.

After receiving all available search information for a turn, A must calculate his "best guess" as to the likelihood that a launcher is at a given site. That is, A must update his belief matrix. A number of approaches can be used to update such beliefs. One such

[16] CNA Research Memorandum (CRM) D0007164.A1, *Using Gaming and Agent Technology to Explore Joint Command and Control Issues*, by Peter P. Perla, Andrew Ilachinski, Carol M. Hawk, Michael C. Markowitz, and Christopher A. Weuve, October 2002. Hereafter cited as Perla et al., 2002.

method is the classic approach of Bayesian updating. Another, the one we employed in this initial work, used the *Durkin Summation* function that is commonly used in fuzzy-logic applications.[17]

Once the agents have updated their belief matrix, they must choose which of the potential target squares to designate for possible strike. We used a simple threshold criterion for making these selections. A's strike plan consists of reading off the top ranking sites and communicating this strike-recommendation to the other agents and game's output routines.

The last major element of an agent's personality has to do with how the agent decides to use his sensor, given that he has just updated his belief matrix for the entire playing field. To design an agent logic that is both flexible enough to encompass a variety of decision "types" (to provide the user with some parametric variability for experimentation) and simple enough to avoid overwhelming the user by the number or complexity of the parameters at his disposal, we considered the basic kinds of motivations that an agent must weigh in deciding where to place his assets. (Some are intrinsic motivations to maximize information gain; others are associated with what an agent presumably knows, or believes, about sensor capabilities.) For example, an agent may choose to maximize the number of squares covered by at least one sensor on the given turn. Another possibility is that the agent will seek to minimize the number of sites that have not yet been searched.

In any case, for a given turn, an agent considers all possible options of placing each of the sensors under his control, and calculates the *Sensor Placement Fitness Function* for each position, based on the set of motivations important to that agent. The form of such fitness functions can be as simple as a weighted sum with fixed weights for each motivation, or as complex as one that varies motivations with time and the game situation.

4.2 THE SOFTWARE IMPLEMENTATION

The software we designed to implement the agent-based *SCUDHunt* model sketched out in the previous section uses an object-oriented architecture that defines the game, the agents, and the assets as objects. The game controls its players, and the players control their assets. Agent beliefs evolve as the game progresses; they are initially defined by parameter values provided by the user in the set-up routine of the database.[18]

Calculation of partial, overall, and cumulative beliefs depends on the communication mode among the agents. To correspond to the modes employed in the human-based games, we implemented two options—the exchange of raw data, and the exchange of current beliefs.

---

[17] The performance of expert systems based on certainty factors has, on occasion, outperformed Bayesian reasoning (at least in systems designed to mimic human diagnostic judgment). See John Durkin, *Expert Systems: Design and Development*, Prentice Hall, 1994.

[18] For more details about the computer model, see Perla et al., 2002.

In the case of raw-data exchange, each agent receives the asset name and search result of every other agent. He interprets these data based on his assessment of each sensor's reliability to arrive at his own partial belief. The agent then incorporates this information into his own overall belief after modifying (multiplying) the partial belief by his own trust in the agent who was the source of the sensor information.

In the case of sharing beliefs (or interpreted data), each agent receives the partial belief of every other agent for each site in the game. Agents then modify those partial beliefs based on their trust in each of the other agents. The agents then incorporate this modified belief into their own overall beliefs.

In our initial implementation, our fitness function used a weighted sum of three motivations: (1) maximize board coverage, (2) emphasize high-belief cells, and (3) de-emphasize cells that have exceeded threshold belief. The function assigned each cell a weight. Agents are more likely to search cells with higher weights.

To control the set-up and execution of the model, and to collect the output data from its use, we implemented a Microsoft Access database application. Information recorded in the database for each game includes the game's initial parameters, such as the number of turns and the communication mode. The user also defines the Agent personalities, including initial assessments of asset reliability and partial beliefs as a function of the various search results possible for each asset. Other necessary parameters include weights for the various asset-placement motivations, and asset assignments to the various agent-players.

The game engine sends the results of each turn and each overall game to the database to be recorded. The data include actual target locations, nominated target locations for each agent, asset placement and search result for each asset for each turn, cumulative belief for each agent over all turns (the final cumulative belief for each cell), the SSA score, and the accuracy score.

5. THE MINI-EXPERIMENTAL CAMPAIGN

5.1 THE HUMAN-BASED EXPERIMENT[19]

The purpose of the July 2002 human-based experiment was to explore how different qualities of information and different types of shared-visualization tools might affect shared situational awareness and accuracy of decisions. The experiment used the same measures of SSA and accuracy described earlier.

As in prior *SCUDHunt* experiments, six four-person teams played the game. Each player managed one or two search assets. Also as in the original *SCUDHunt* experiment, the statistical experimental design used a Latin Square with factorial treatments. In total,

[19] For a complete discussion of the human-based experiment, see TLI, 2002.

there were six treatment combinations, defined by three levels of quality of information (QOI), high, medium, and low, and two types of visualization techniques, shared visualization (shared vis) and post visualization (post vis). In addition, players could communicate directly using text chat in all games.

The Naval War College provided the players for this game, primarily from naval Reservists doing a summer tour in Newport. They also provided the facilities for the players to conduct the game.

Each of six teams played one game with each treatment combination. To control for the likely team effects and the possible effects of learning over the course of the six-game set, the order in which each team played the different treatment combinations was different, as shown in table 2. The treatments themselves are defined below.

Table 2: Latin Square Design

	Game					
Team	1	2	3	4	5	6
1	B	E	A	C	F	D
2	D	A	E	B	C	F
3	E	B	C	F	D	A
4	A	F	D	E	B	C
5	F	C	B	D	A	E
6	C	D	F	A	E	B

A: QOI High, Shared viz — C: QOI Med, Shared viz — E: QOI Low, Shared viz
B: QOI High, Post viz — D: QOI Med, Post viz — F: QOI Low, Post viz

We defined three levels of information quality. Practically speaking, these three levels were defined by the set of probabilities of various reports from the different sensors.

- QOI Medium—Probabilities are the same as in all prior *SCUDHunt* experiments. This represents the base case. There is a small chance of false positives and false negatives.

- QOI High—False-negative results, defined as a result of 0 returned from the search of a square containing a launcher, are decreased. Three assets that in the base case were likely to return a 0 result in a square containing a launcher, return a ? in this case.

- QOI Low. False positive results, defined as a result of X returned from the search of an empty square, are increased.

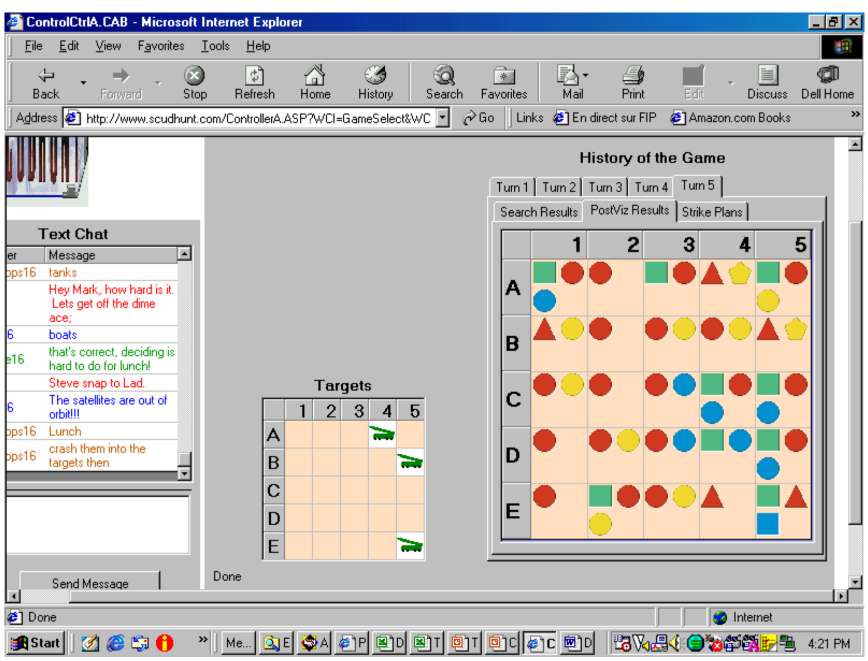
The second factorial treatment introduced two types of visualization techniques to help the players track the course and results of their search efforts. Half the games used what we called shared-visualization and the other half used post-visualization.

The original *SCUDHunt* implementation included a graphical display that allowed players to see the search results displayed on an image of the search space. We subsequently modified and improved this display. After the results of each player's searches were calculated, the game displayed these results both by individual sensor and in a combined display for each turn of the game. In the latter case, the identity of the sensor is indicated by the color of the background of the symbol.

The post-visualization tool was originally designed by the Naval War College for their earlier experiment. The post-visualization tool is similar in structure to the shared-visualization one, but instead of the game program's reporting the results of each search automatically, in post visualization the process requires two steps. First, players receive a depiction of the search results returned by their own assets. Players then are given the capability to insert pre-defined symbols, designed by the Naval War College, on a copy of the game board similar to that used in the shared-visualization case. These symbols reflect the identity of the player who uses it and different degrees of certainty about the presence or absence of SCUD launchers in each square on the board. Players could place symbols on as many squares as they wished.

The levels of certainty are defined qualitatively as: No Information, No SCUD, Possible SCUD, Probable SCUD, and Confirmed SCUD. Once each player has "posted" his "belief values" into the post-visualization system, the program presents an aggregated picture to all the players. As in the shared-visualization tool, the results are color-coded by player and there is a tab for each turn, so players can review results from previous turns.

Figure 5: Post-visualization display



5.2 THE AGENT-BASED EXPERIMENT[20]

To compare the human-based experiment to an agent-based one, we defined agent-based analogs to the experimental treatments. This required us to develop a range of values for certain of the agent parameters, particularly the *Sensor Report-Launcher Correlation* matrices and the *Trust* matrices to distinguish each of the 24 agents from the others, and to reflect the experimental conditions. We then ran a series of 36 agent-based games using the same experimental design as the human experiment (although in this case we made no attempt to introduce any effect for the actual sequence of games played by each of our agent teams). We created six teams of four agents each to play the required series of six games per team.

The three levels of information quality in the experiment required us to reflect how the agents should play the game differently as a function of information quality. The human players were informed of the nature of the information they were to receive in each of their games (although they were not given the actual probabilities of different search outcomes). We thus defined the agent characteristics for each of the three levels of information quality by the values of the *Sensor Report-Launcher Correlation* matrix for each player for each case.

To do this, we decided to use a baseline of three values for each agent for each possible outcome for each sensor—high, medium and low. A high result, for example, meant that the agent would have a strong belief that the sensor was providing an accurate indication of the actual state of the searched location. A low value meant that the agent would be less certain of an accurate result, thus creating a smaller partial belief value. These values were modified for the three variants of information quality to reflect the differences in how agents might interpret the sensor results based on those different levels. We then defined our 24 game-playing agents by randomly selecting one of the three values for each of the required parameters from our previously defined set for each of the three levels of information quality.

The two visualization techniques used in the human games are more easily represented in the agent-based model. The shared visualization treatment is analogous to allowing the agents to share only raw data. The post visualization treatment is analogous to allowing the agents to share only their current beliefs.

In both cases, however, it is necessary to distinguish each agent from the others based on their personality elements, or all of the agents would develop identical beliefs based on the shared data. In the case of sharing raw data, we relied on the differences in the *Sensor Report-Launcher Correlation* matrices of the agents to make this distinction. In the case of sharing beliefs, however, we introduced specific values for the *Trust* matrices of the players. We used the same general approach to defining the values of the *Trust* matrices as we did for the *Sensor Report-Launcher Correlation* matrices. Each agent's trust of each other agent was chosen at random from among three possible trust values.

---

[20] For a complete discussion of the agent-based experiment, see Perla et al., 2002.

In addition to the parameter definitions described above, we gave each agent different values for the necessary threshold parameters.

To define the strike recommendations and SSA scores for the agent-based games, we required that each agent nominate a minimum of three target locations, even if their nomination threshold would normally prevent them from doing so. If an agent did not automatically nominate the required minimum, we simply chose the locations with that agent's three highest final belief values, along with any other locations that had belief values the same as the lowest of the three.

5.3 COMPARING HUMAN- AND AGENT-BASED RESULTS

This experiment provided us with a first opportunity to explore the practicality of creating agent-based experiments that are analogous to human-played ones. We did not have the time and resources to do an extremely detailed comparison of both results and processes of play in both experiments. Our comparison is, therefore, limited to an initial look at the overall outcomes of the two experiments in terms of SSA and accuracy, both at the level of individual game scores and the overall ANOVA.

Table 3 shows the data for the SSA scores of the agent-based experiment (treatment codes are shown in parentheses). Compare those results with the ones shown in table 4, for the human-based experiment. A cursory examination of the raw data of the human-based game indicates that team 1's scores tend to be noticeably lower than the others, and that those of teams 2 and 6 seem to be higher, with the other three teams falling in the middle of the range. The agent-based SSA scores are higher on average (3.52) than those of the human-based game (2.74).

Table 3: SSA scores, agent-based experiment

Team	Game 1	Game 2	Game 3	Game 4	Game 5	Game 6	Mean
1	3.25 (B)	3.40 (E)	4.00 (A)	4.00 (C)	3.00 (F)	4.00 (D)	3.61
2	3.40 (D)	3.50 (A)	3.50 (E)	2.40 (B)	4.00(C)	2.67 (F)	3.24
3	3.12 (A)	4.00 (F)	4.00 (D)	2.67 (E)	3.40 (B)	3.50 (C)	3.45
4	4.00 (F)	4.00 (C)	4.00 (B)	3.12 (D)	3.75 (A)	3.75 (E)	3.77
5	3.25 (E)	2.17 (B)	3.25 (C)	4.00 (F)	4.00 (D)	3.50 (A)	3.36
6	32.60 (C)	4.00 (D)	4.00 (F)	4.00 (A)	3.86 (E)	3.850 (B)	3.66

Treatments A: QOI High, Shared viz C: QOI Med, Shared viz E: QOI Low, Shared viz
B: QOI High, Post viz D: QOI Med, Post viz F: QOI Low, Post viz

Table 4: SSA scores, human-based experiment

Team	Game 1	Game 2	Game 3	Game 4	Game 5	Game 6	Mean
1	2.09 (B)	1.89 (E)	1.92 (A)	2.00 (C)	1.82 (F)	2.15 (D)	1.98
2	2.00 (D)	3.00 (A)	4.00 (E)	4.00 (B)	4.00 (C)	4.00 (F)	3.50
3	2.50 (A)	4.00 (F)	2.60 (D)	1.86 (E)	1.82 (B)	1.82 (C)	2.43
4	2.60 (F)	3.00 (C)	2.56 (B)	2.00 (D)	3.50 (A)	1.82 (E)	2.58
5	3.00 (E)	2.40 (B)	2.17 (C)	3.00 (F)	2.40 (D)	1.88 (A)	2.47
6	4.00 (C)	3.25 (D)	3.20 (F)	4.00 (A)	2.50 (E)	4.00 (B)	3.49

Treatments A: QOI High, Shared viz C: QOI Med, Shared viz E: QOI Low, Shared viz
B: QOI High, Post viz D: QOI Med, Post viz F: QOI Low, Post viz

The agent-based model may help us better reflect the complexities of differences in human belief systems and trust for each other's judgments, but at this early stage of development, we have not been able to vary parameter values to explore those dynamics.

Table 5 shows the ANOVA table for the SSA scores of the agent-based experiment. Compare those results with the ones shown in table 6, for the human-based experiment. These tables show the analysis of variance for the effects of team, game, and treatments. The treatments effect is further decomposed into effects for the two crossed factors, quality of information (QOI) and visualization, separately, along with their interaction. The QOI factor is further decomposed by two orthogonal contrasts, the first between the average of medium and low levels of information and the high level, and the second between the medium and low levels.

The rightmost column of p-values indicates the statistical significance of the results. These p-values can range from 0 to 1. The lower the p-value, the stronger the indication of a significant effect for that factor. Traditional p-values for "statistically significant" results range between 0.01 and 0.05.

On that basis, the only significant effects we can observe in these data are those exhibited by the teams—not surprising, given our earlier observation of their apparent differences in the raw data. However, there is no real evidence that any of the treatment combinations have a measurable effect on SSA. Furthermore, it is particularly gratifying in this case to see that the p-value associated with the order of a game in the sequence (the column effect) shows little evidence that a learning effect has muddied our more substantive explorations. This result could be the effect of a pre-game training program that each of the players received.

Table 5: ANOVA of SSA scores for agent-based experiment

	Sum of squares	Degrees of freedom	Mean square	F statistic	p-value
Team (row)	1.18	5	0.24	0.83	0.54
Game (column)	1.10	5	0.22	0.77	0.58
Treatment	1.73	5	0.35	1.22	0.34
QOI	0.55	2	0.27	0.97	0.40
Med/Low - High	0.12	1	0.02	0.41	0.53
Med - Low	0.43	1	0.001	1.53	0.23
Visualization	0.06	1	0.06	0.22	0.64
Interaction	1.12	2	0.56	1.98	0.16
Error	5.67	20	0.28		
Total	9.68	35			

Table 6: ANOVA of SSA scores for human-based experiment

	Sum of squares	Degrees of freedom	Mean square	F statistic	p-value
Team (row)	11.48	5	2.30	5.25	0.003
Game (column)	0.37	5	0.07	0.17	0.97
Treatment	2.90	5	0.58	1.33	0.29
QOI	0.84	2	0.42	0.96	0.40
Med/Low - High	0.84	1	0.84	1.91	0.18
Med - Low	0.0001	1	0.0001	0.0002	0.99
Visualization	0.46	1	0.46	1.06	0.32
Interaction	1.60	2	0.80	1.83	0.19
Error	8.75	20	0.44		
Total	23.50	35			

Consistent with the raw numbers, we see that the agent-based game exhibited no significant team effect. Again, this is not surprising given the fact that the teams were created using the same randomization technique and so we would not expect them to exhibit any significant differences.

Table 7 presents the data for the accuracy scores of the agent-based experiment. Compare the results with those in table 8, the human-based results.

Table 7: Accuracy scores, agent-based experiment

Team	Game 1	Game 2	Game 3	Game 4	Game 5	Game 6	Mean
1	0.38 (B)	0.41 (E)	0.33 (A)	0.67 (C)	0.33 (F)	0.67 (D)	0.47
2	0.65 (D)	0.38 (A)	0.38 (E)	0.75 (B)	0.33 (C)	0.06(F)	0.42
3	0.32 (A)	1.00 (F)	0.40 (D)	0.00 (E)	0.31 (B)	0.43 (C)	0.41
4	1.00 (F)	0.50 (C)	0.67 (B)	0.32 (D)	0.73 (A)	0.47 (E)	0.61
5	0.69 (E)	0.38 (B)	0.31(C)	0.25 (F)	0.67 (D)	0.38 (A)	0.45
6	0.62 (C)	0.67 (D)	0.00 (F)	0.67 (A)	0.30 (E)	0.14 (B)	0.40

Treatments A: QOI High, Shared viz C: QOI Med, Shared viz E: QOI Low, Shared viz
B: QOI High, Post viz D: QOI Med, Post viz F: QOI Low, Post viz

Table 8: Accuracy scores, human-based experiment

Team	Game 1	Game 2	Game 3	Game 4	Game 5	Game 6	Mean
1	0.35 (B)	0.41 (E)	0.44 (A)	0.67 (C)	0.5 (F)	0.43 (D)	0.47
2	0.67 (D)	0.92 (A)	0.67 (E)	1.00 (B)	1.00 (C)	0.67 (F)	0.82
3	0.15 (A)	1.00 (F)	0.62 (D)	0.23 (E)	0.35 (B)	0.60 (C)	0.49
4	0.62 (F)	0.80 (C)	0.39 (B)	0.67 (D)	0.86 (A)	0.45 (E)	0.63
5	0.33 (E)	0.67(B)	0.77 (C)	0.83 (F)	0.75 (D)	0.27 (A)	0.60
6	1.00 (C)	0.62 (D)	0.50 (F)	1.00 (A)	0.50 (E)	1.00 (B)	0.77

Treatments A: QOI High, Shared viz C: QOI Med, Shared viz E: QOI Low, Shared viz
B: QOI High, Post viz D: QOI Med, Post viz F: QOI Low, Post viz

Once again the raw data for the human game seem to indicate the same partitioning of the teams that we observed in the case of SSA: team 1 seems to score a bit lower than the others (though team 3's scores are not much better), while teams 2 and 6 seem noticeably better than the others. In this case, the raw data from the agent games seems reasonably consistent with the performance of the lower or average human teams, but none of the agent teams really approaches the performance of the best of the human teams. Of note, however, is the fact that we see a number of specific games in which the accuracy performance of the agents is essentially identical to that of their corresponding human teams. For example, team 1's second game had an accuracy score of 0.41, and team 4's first game had an accuracy score of 1.00 and an SSA score of 4.00. We should not make too much of this coincidence in the scores, but it is encouraging to see such similarities at such an early stage of development of the agent-based model.

Tables 9 and 10 are the ANOVA tables for the data above.

Table 9: ANOVA of accuracy scores for agent-based experiment

	Sum of squares	Degrees of freedom	Mean square	F statistic	p-value
Team (row)	0.19	5	0.04	0.73	0.61
Game (column)	0.33	5	0.07	1.027	0.31
Treatment	0.47	5	0.09	1.81	0.16
QOI	0.44	2	0.22	4.20	0.03
Med/Low - High	0.20	1	0.20	3.89	0.06
Med - Low	0.24	1	0.24	4.50	0.05
Visualization	0.003	1	0.003	0.06	0.80
Interaction	0.03	2	0.02	0.30	0.74
Error	1.04	20	0.05		
Total	2.04	35			

Table 10: ANOVA of accuracy scores for human-based experiment

	Sum of squares	Degrees of freedom	Mean square	F statistic	p-value
Team (row)	0.62	5	0.12	4.58	0.006
Game (column)	0.26	5	0.05	1.93	0.14
Treatment	0.69	5	0.14	5.17	0.003
QOI	0.54	2	0.27	10.12	0.0009
Med/Low - High	0.38	1	0.38	14.09	0.001
Med - Low	0.16	1	0.16	6.16	0.02
Visualization	0.0002	1	0.0002	0.009	0.92
Interaction	0.15	2	0.075	2.80	0.08
Error	0.54	20	0.03		
Total	2.10	35			

The observation of differences among human teams is borne out by the statistical analysis. The p-value for teams is again quite low, at 0.006. In addition to the team effect, this time we see several other statistically significant results. The treatments as a whole show a strong p-value of 0.003. Our decomposition of the treatments shows that the significant effects seem to reside primarily in the effect of information quality, surely not a surprise. Once again, the effect of the different visualization techniques shows no evidence of significance. Also as we saw above, the team effect detected in the human experiment is absent from the agent experiment. Both experiments agree on the significance of information quality as an important effect on the accuracy of decisions, but the agent-based experiment does not produce quite as strong a body of evidence (a p-value of 0.03 compared to the 0.0009 value of the human experiment.)

5.4 SPECULATION ON THE POTENTIAL OF THE AGENT-BASED SYSTEM

The results presented in the preceding section are mere examples of the potential value of using the agent-based model as a research tool in conjunction with human

experiments. We were unable to pursue further the comparative analysis between the two experiments because of time and resource constraints. But some of the possible directions of such exploration are apparent.

First, it is clear that the method we used to define agent personalities and team composition produced similar team characteristics, unlike the more disparate teams we see in the human experiment. What characteristics of individual players or teams might have contributed to the generally better performance of teams 2 and 6 in the human experiment? Did those players perhaps have a better *Sensor Report-Launcher Correlation* matrix than the other teams? Did they know each other better and thus trust each other more than the others?

We can explore some of these issues using the agent-based model. To do so we would need to explore the value space of the various parameters we used to define the agents and their team interactions. For example, we can change the values of the *Sensor Report-Launcher Correlation* matrices of the players of two teams to make them both more accurate and more similar to each other than those of the other teams. How does such a change affect SSA and accuracy scores for those teams and the overall ANOVA for the experiment? Similar explorations of issues associated with the *Trust* matrices might help us investigate those effects on team performance.

In addition to gross output measures, it is also possible to explore the internal dynamics of how the agents conduct their searches. The model records where each agent places each asset during each turn of a game. A detailed comparison of these data with the corresponding data from the analogous human-based game may lead to new insights about the differences in the dynamics of how agents and humans actually make decisions about asset placement. This could help us refine our fitness functions to make them better reflect the decision logic we have observed in the human players.

Some six years ago, the defense community began to notice the initial research into the adaptation of complex-systems theory to combat.[21] Since that time, each year has seen new developments in the field, just as the broader subjects of non-linear dynamics, cellular automata, and complexity theory make more and more inroads into the way we think of science in general.

Our initial attempt to study command and control using an agent-based approach—based on the *SCUDHunt* experimental testbed—has taken only a first step along what we believe may be a similar path toward the development of new techniques with which to study these important issues.

---

[21] See, for example, the following CNA papers by Andrew Ilachinski: CNA Information Memorandum (CIM) 461.10, *Land Warfare and Complexity, Part I: Mathematical Background and Technical Sourcebook, First Revision*, July 1996; CNA Research Memorandum (CRM) 96-68, *Land Warfare and Complexity, Part II: An Assessment of the Applicability of Nonlinear Dynamics and Complex Systems Theory to the Study of Land Warfare*, July 1996; CNA Research Memorandum (CRM) 97-61.10, *Irreducible Semi-Autonomous Adaptive Combat (ISAAC): An Artificial-Life Approach to Land Warfare, First Revision*, August 1997; CNA Annotated Briefing (CAB) 97-88, *A Concise User's Guide to ISAAC-FL: ISAAC's Mission-Fitness Landscape Mapper Program*, September 1997; and CNA Research Memorandum (CRM) D0007376.A1, *Multiagent-Based Synthetic Warfare: Toward Developing a General Axiological Ontology of Complex Adaptive Systems*, January 2003.

6. THE WAY AHEAD

Web-based technologies, specifically distributed games, offer many benefits to analysis and training—they can save a lot of money, and occasionally provide a more realistic environment than is available using traditional models and simulations. Traditional models and simulations often overlook the analysis of such "soft factors" as building trust in virtual teams, learning how to communicate with individuals and organizations with different cultures, understanding their capabilities and resources, and building a shared picture or SSA. These factors can be taught and analyzed in on-line, distributed environments. In addition, games of all types are particularly useful for exploring cooperation, coordination, communication, risk taking, problem solving, leadership, group dynamics, and team building. This report highlights how distillation games can create powerful analytical environments. The *SCUDHunt* experiments successfully showcased the effectiveness of Internet-mediated games as analysis tools for studying complex problems. One particular advantage of Internet-based games is that they can be instrumented and their results directly mapped to the experiment's design variables and outcomes.

SCUDHunt in particular is well suited to experiments focused on information sharing, information quality, new warfighting concepts, decision-making and new command and control strategies. The game can be made more or less complex to suit the underlying research agenda. For instance, the enemy is stationary in the current version of *SCUDHunt*. Future versions of the game could include an enemy that maneuvers in the battle space and employs decoys; we could also model sensors whose reliability varies over time (e.g. a spy is "turned" to give faulty reports).

The use of adaptive agents within the context of game-based experimentation can help address one of the main difficulties of experimentation with human players—finding appropriate numbers and types of human players for the game. Using agent-based gaming will allow us to explore the experimental design space more thoroughly and much more quickly than possible using games with live participants. Such wide-ranging analysis can help us focus precious human experimentation on issues with the greatest potential payoff.

As we look to the future, we are struck by today's current rage for "transformation." DoD has established an office whose primary purpose is to advocate and pursue the transformation of the U.S. military establishment. Panels and study groups are convened and meet to report on whether new ideas are, or are not, transformational enough to be considered for future funding.

To transform the way we act, however, we must first transform the way we think. And in the world of command and control, much of the way we think is bound up in how we define problems analytically, how we conduct exercises and experiments, how we collect data, and how we assess the data for whatever evidence we can find to help us assess the practical value of new ideas and equipment. Our research has convinced us that, at the very least, we must transform our thinking about how to study and evaluate military command and control in two specific dimensions.

First, we must integrate the new sciences of agent-based modeling and the study of emergent phenomena into the existing techniques for understanding command-and-control issues. The same complexities and non-linear behaviors that form the basis for applying similar techniques to the study of combat dynamics are inherent as well in the command and control of such combat.

Second, we must integrate game-based experimentation using distillation games into the existing routine of demonstrations, experiments, and exercises through which the C2 community of DoD currently seeks to explore future concepts of military command and control. Demonstrations tend to focus on component elements of the very complex systems of people, procedures, and equipment that make up the military C2 system. Often experiments are so large and costly to put on that failure is not an option and learning the truth becomes more of an obstacle than an opportunity. Exercises suffer from this same attitude to an even greater extent, if the controversies surrounding Millennium Challenge 2002 are any indicators.

Game-based experimentation is a scientifically based and statistically valid technique that can help us explore practical questions about human performance in C2-related tasks. Such insights are of fundamental importance if we are to improve our understanding and representations of such operational concepts as network-centric warfare, information warfare, and self-synchronizing command systems. Modeling the interactions inherent to these concepts, and testing hypotheses about key factors are critical to making sustainable scientific progress in this field. As the research based on *SCUDHunt* has shown, game-based experimentation offers definite promise in this area.

The *SCUDHunt* "universe" is simple enough to allow us to conduct a comprehensive exploration of its dynamics, and yet rich enough that such an exploration provides useful insights into real-world issues. Even the basic research reported here indicates that our agent-based system can recreate some of the elements of human behavior to a useful level of fidelity. As we enhance our understanding and refine our model, we can see whether some universal traits and behaviors begin to emerge across a spectrum of agent types.

In order to pursue the promise suggested by our results, it is necessary to develop some basic research tools and approaches for using both the human and agent versions of *SCUDHunt*, and especially for integrating both versions into a unified research program. As the foundation for this research program, we should first systematically verify the results we have already obtained in human games. We should develop different pools of agents of various types and conduct experiments to determine whether we see outcomes influenced by the same sorts of communications topologies or behavioral variables that we see reflected in the human games and experiments.

Building on this foundation, we can define a set of "basis" agents, or archetypes, which we can combine in various ways to span the full set of agent behaviors. Such basis agents may reflect directly the different dimensions of agent personality we have already defined (for example, the "trusting" agent, who believes everything everyone tells him, or the "skeptic" agent, who believes nothing). In any case, such basis agents should, to the extent possible, reflect our understanding and intuition of obviously different kinds of agents.

Using the basis agents as a starting point, we can develop genetic algorithms or similar techniques to sweep out wide subsets of the parameter spaces of greatest interest to practical problems. If we can identify a dynamic gestalt emerging from such experimentation that is similar to what we have seen and documented in human performance, it can lend credibility to whatever drivers we believe we have identified in the human experiments. If for some reason we cannot identify such similarities, the reasons that we cannot may shed light on:

- What we need to change or improve about the behaviors programmed into our agents
- Something interesting and entirely different and surprising.

Such is the nature of the process of using complex-systems models to explore complex real-world processes.

SELECTED RELATED PUBLICATIONS

Exploring Joint Force Command and Control Concepts Using SCUDHunt – Final Report, Marcy Stahl, Julia J. Loughran, ThoughtLink, Inc., Joint C4ISR Decision Support Center, October 2002

This document describes the ThoughtLink 2002 experiment using *SCUDHunt* to explore the effect of three levels of quality of information and two kinds of visualization on shared situational awareness and accuracy of decisions.

Meta-Analysis of Team Performance Accuracy and Shared Situational Awareness in SCUDHunt Experiment, Jim Holzworth, Paradigm Associates, Joint C4ISR Decision Support Center, October 2002

This meta-analysis identifies statistical relationships between various factors across four experiments: the ThoughtLink 2002 experiment, the ARI 2002 experiment, the NWC 2001 experiment, and the original ThoughtLink/CNA 2000 experiment.

C2 Concepts and Experimentation Literature Review, Rebecca Agrait, Julia Loughran, Colleen Mackey, Marcy Stahl, ThoughtLink, Inc., Joint C4ISR Decision Support Center, October 2002

ThoughtLink reviewed academic and military research in several areas: shared situational awareness, information sharing, and team structure. Interesting results and some metrics are described in this review.

Key Drivers for C2 Performance: Data Mining SCUDHunt Experiment Data, Julia Loughran, Marcy Stahl, Peter Perla, ThoughtLink, Inc., Joint C4ISR Decision Support Center, November 2001

We did some additional analysis of the TLI/CNA 2000 experiment, focusing on accuracy of players' decisions and regression analysis of player backgrounds and subjective assessments and their relationship to shared situational awareness.

SCUDHunt Data Dictionary, ThoughtLink, Inc., Joint C4ISR Decision Support Center, September 2001

This describes the format and type of data saved during *SCUDHunt* execution. These data are then used in subsequent analyses. (Since this dictionary was written, we've added two new tables to accommodate another visualization capability.)

Gaming and Shared Situation Awareness, Peter P. Perla, Michael Markowitz, Albert Nofi, Christopher Weuve, Center for Naval Analyses; Julia Loughran, Marcy Stahl, ThoughtLink; DARPA, November 2000

This CNA document describes the first *SCUDHunt* experiment in 2000, in which three modes of communication and the availability of a shared visualization were varied.

Defining and Measuring Shared Situational Awareness, Albert Nofi, Center for Naval Analyses; DARPA, November 2000

This CNA document discusses different formulations of shared situational awareness.

The DICE Experiment: Creating and Evaluating a Web-based Collaboration Environment for Interagency Training, ThoughtLink, Inc., DARPA and C4ISR Cooperative Research Program, May 2000

ThoughtLink created a prototype collaborative environment for distributed teams and then conducted an experiment comparing the effectiveness of face-to-face training with the distributed training.

Applying Commercial Gaming and Collaboration Technologies to Augment JTF Staff Training, ThoughtLink, Inc., DARPA, January 1999

This report reviews various simulations used by the military for Joint Task Force staff training and then describes how ideas from commercial gaming and collaboration technologies could be used to create a distributed collaborative environment for more frequent, lower fidelity, lower cost training.

Using Gaming and Agent Technology to Explore C2

17 June 2003

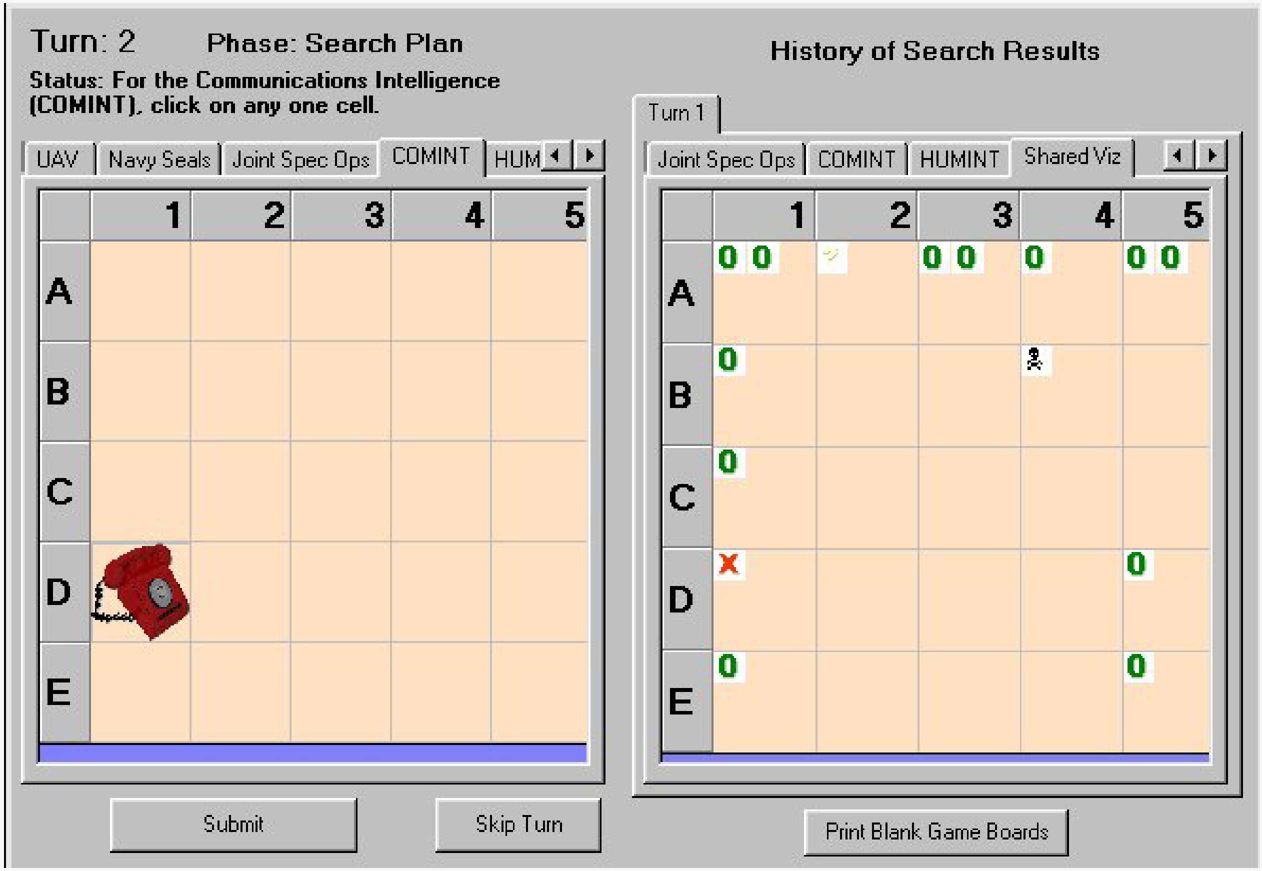
Presentation to
8th ICCRTS

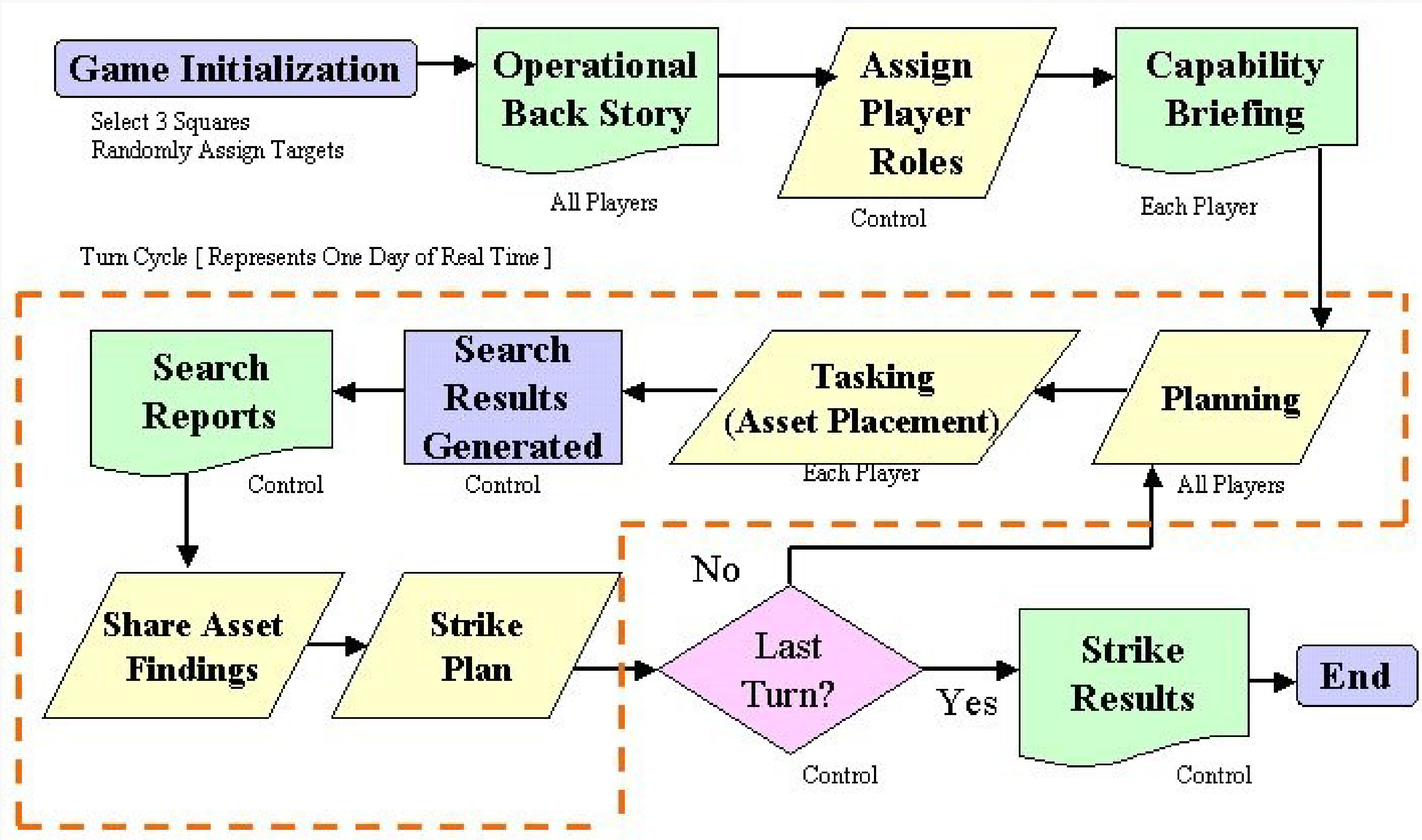
Presented by:
Julia Loughran & Peter Perla

Summary

- We can create distillation games that capture the key elements in the OODA loop
- We can use such games to create experiments that are amenable to statistical design and analysis
- We can use game-playing agents and genetic algorithms to explore vast C2 decision spaces
- We can use human games to validate findings, suggest adjustments, and identify new areas for exploration
- We can integrate agent and human games in experimental campaigns to address fundamental issues systematically

SCUDHunt sample gameboard





Game Initialization
Select 3 Squares
Randomly Assign Targets
Operational Back Story
All Players
Assign Player Roles
Control
Capability Briefing
Each Player
Turn Cycle [Represents One Day of Real Time]
Search Reports
Control
Search Results Generated
Control
Tasking (Asset Placement)
Each Player
Planning
All Players
Share Asset Findings
Strike Plan
No
Last Turn?
Control
Yes
Strike Results
Control
End

Experimental measurements

Shared Situational Awareness (SSA) score- overlap in assessment of launcher locations among team members, irrespective of whether understanding is right or wrong

SSA score = Ratio of the total number of recommended target squares by all players to total number of unique squares designated

Example: Perfect SSA: All 4 team members vote for the same 3 squares = 12/3 = 4
Lowest score: All 4 team members vote for 3 different squares = 12/12 = 1

Experimental measurements

Accuracy (ACC) score - Do team members (or individual players) actually find the launchers?

ACC = ratio of nominated squares that actually contained SCUD launchers to the total number squares nominated

Example: Perfect team ACC: 4 players vote for the same 3 squares containing launchers = 12/12 = 1.

Lowest ACC: Team does not identify any launcher squares, then their score is 0 / 12 (or some other large number) = 0

We also compute individual player ACC

Experiment/Year	Conducted by	For	Experimental Variables
Experiment #1; 2000	ThoughtLink and CNA	DARPA	Availability of visualization, type of communication
Data Mining of Experiment #1; 2001	ThoughtLink	Joint C4ISR Decision Support Center	Data mining of original experiment for quality of decisions
Experiment #2; 2002	George Mason University	Army Research Institute	Training on own or all assets, mode of communication
Experiment #3; 2002	Naval War College, CNA, ThoughtLink	Naval War College	Command method, type of visualization
Experiment #4; 2002	ThoughtLink, Naval War College, CNA	Joint C4ISR Decision Support Center	Quality of information, type of visualization
Experiment Meta-Analysis; 2002	ThoughtLink	Joint C4ISR Decision Support Center	Meta Analysis of four *SCUDHunt* experiments

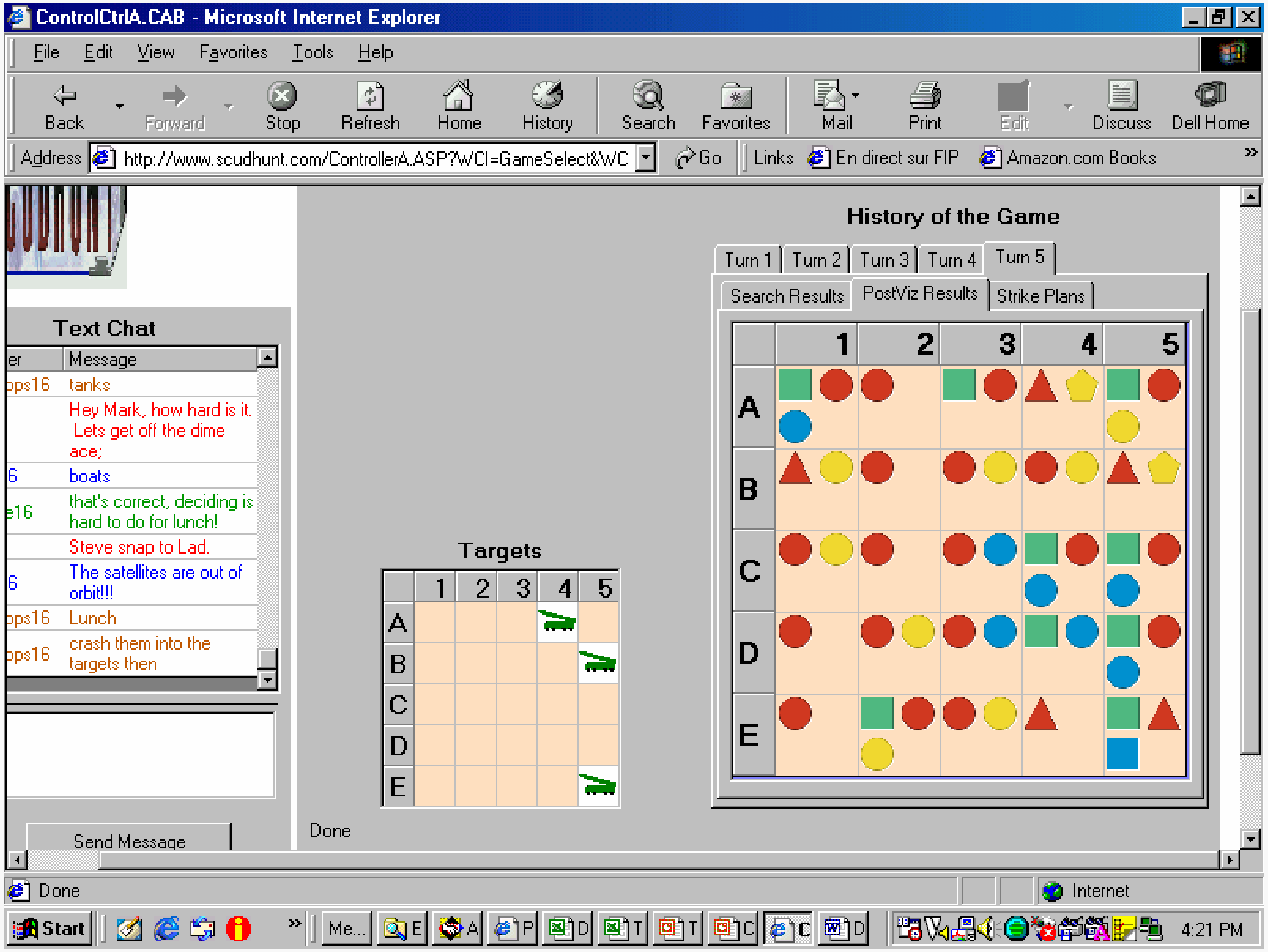
Key results from human experiments

- Quality of information affects ACC more than it affects SSA
  - SSA can be built on bad info, so providing COP is not a cure-all
- ACC and SSA are related
  - From meta-analysis, 50% of variance in ACC can be accounted for by knowing SSA
- Communication matters, but mode of communications doesn’t
  - Chat/voice/shared visualization were similar, in terms of effect on SSA
- What doesn’t matter
  - Duration of games
  - Amount of text chat

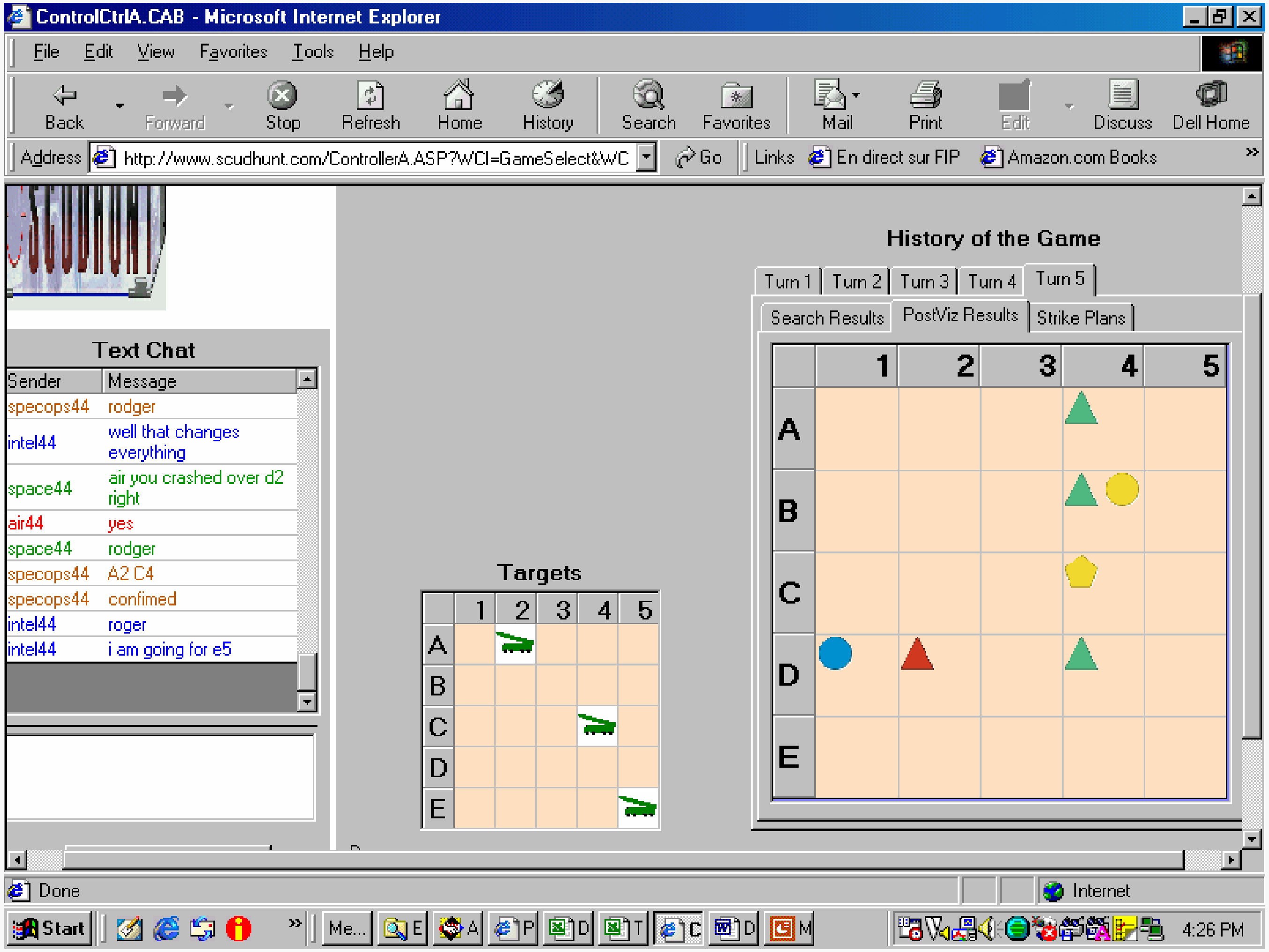
Key results from human experiments

- Teams matter but we’re not sure what is most important
- Teams differ in:
  - Understanding that asset reliability descriptions were critical to success
  - Value placed on timeliness vs. accuracy
  - Degree of integration of their team strategy
  - Leadership style

Team 1's final post viz game - turn 5



Team 4's last post viz game - turn 5



Sample team 3 chat – DSC 2002

Player ID	Message
space35	SPACEto col. 3.
specops35	with your assets up to the ne, I can send the seals across to D2 and joint spec ops up to d5
specops35	Both spec ops will be within search range of E3/E4
air35	maybe spec ops can clear out row E. I'll take manned air over row A and the uav down col 4 so that next space pass will give us corroboration
specops35	I could send the seals down to E2 vs D2 next, but both air and space had e2 clean
specops35	Air, are you thinking Joint Spec ops to E4 this round vs D5
space35	What is level of conf that INTEL is right about E5 (that SEALs chickened out?)?
air35	yes, because you can always move to D5 on a diagonal, right?
intel35	Where is JOint Spec ops starting from? Can they do E4 this turn and D5 next?
intel35	Comint is VERY good at saying a space is Clear
specops35	Yes to intel, they start back in E5. the Koronans ability to hide scuds is low, I think Joint Spec Ops hit that low probability of koronan security with no scud.
specops35	So seals to D2, Joint Spec Ops to E4 this rnd.
space35	Concur.
air35	sounds good
intel35	6 of one, half dozen of another

Sample team 4 chat – DSC 2002

Player ID	Message
intel41	spec ops check out a3
specops41	i am going to check out B2 and E5
intel41	humint checking out b2
space41	which row you guys want me
space41	i'll check row 4
intel41	no scud in b2
space41	4 poss in row 4
specops41	ok i am in a3 and e4
air41	uav killed on e5
specops41	no info in either
intel41	disregard my prob scud in a3 then, my bad
specops41	final go
intel41	this is the search plan that counts
specops41	we know that there is one in C2
intel41	prob in d5, but not sure

Team 2 chat – NWC 2002

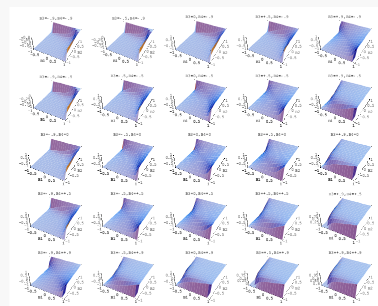
Game	Player	Chat Message
401	intel25	what areas do we not have covered this turn?
401	air25	i dont know
401	specops25	I'll check out B3 but I think intel was there already, only place checked once though
401	air25	just slap joint in somewhere and we will hope that we made good decisions

Team 5 chat - last turn – NWC 2002

Seq-uence	Player	Chat Message
8224	Space56	Definite Negatives: A1, A3, A4, A5, B1,2,3,4,C2,3,4,D2,4,E2,4
8225	Space56	a3 and a4...both nothing
8226	Space56	Probable Negatives (3 no): C1, D1, E1, E3, (2 no): D3
8227	Space56	Mixed: D5 (1 pos, 3 neg)

Integrating human and agent games

Agent Game

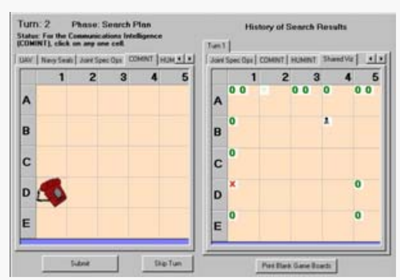


Interesting patterns

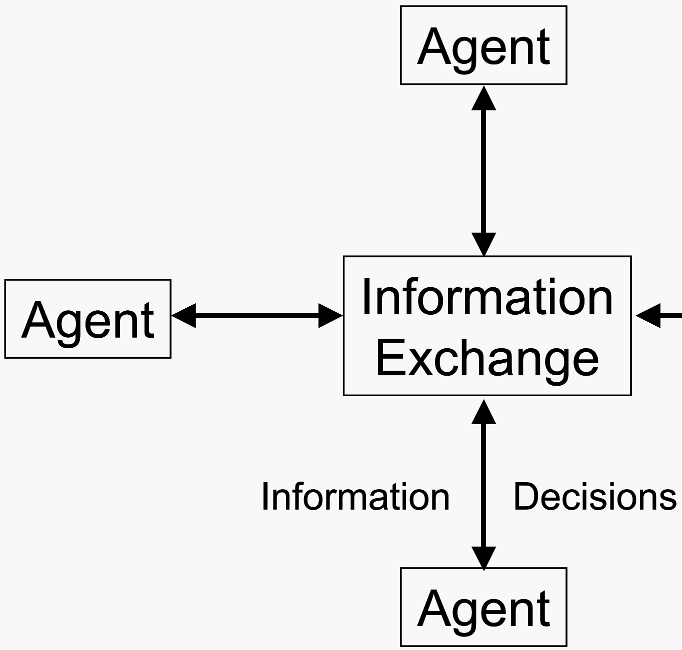
Can we explore why?

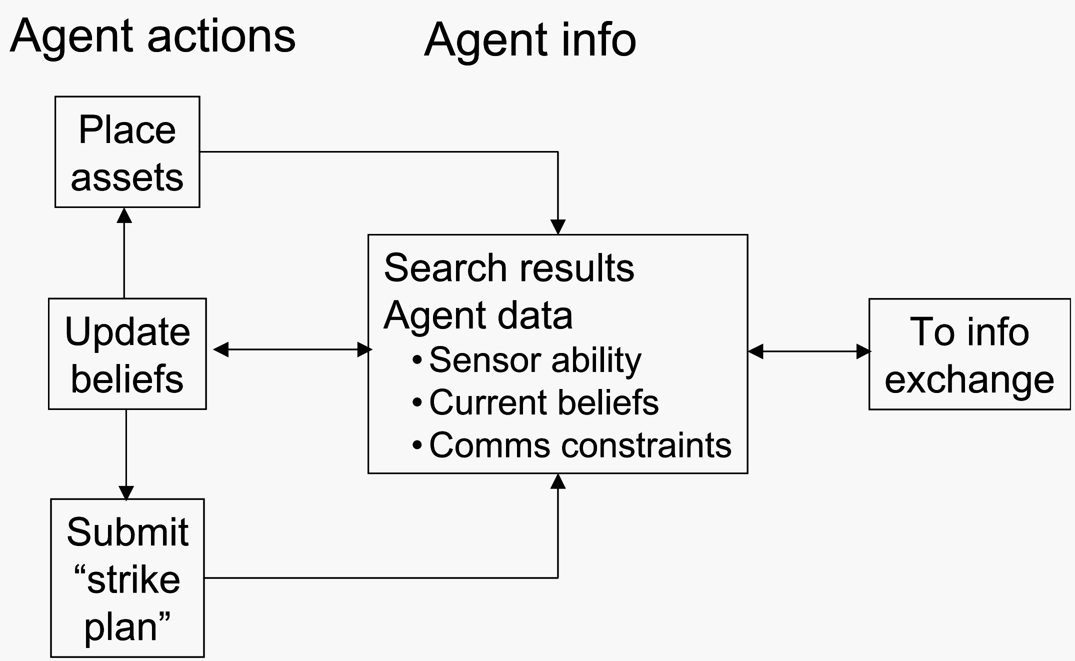
Do humans act that way?

How human players act



Human Game


Agent
Agent
Information Exchange
Agent
Information
Decisions
Agent


Agent actions
Agent info
Place assets
Update beliefs
Submit "strike plan"
Search results
Agent data
• Sensor ability
• Current beliefs
• Comms constraints
To info exchange

The key components of the model include representations of each agent’s:

- *Belief Matrix*, the strength of the agent’s belief that a target is, or is not, present in a specific grid square
- Interpretation of sensor reports and how they change his belief value for the grid squares
- Trust of other agents and how that affects the way he integrates the information they provide into his own belief calculations
- Strike-plan logic, the determination of which targets to recommend for strike
- Sensor-placement logic, the process of deciding where to place the agent’s sensors to maximize some “fitness function” representing the various, possibly competing, motivations an agent may have as he decides how to allocate his search effort.

SSA scores, human-based experiment

Team	Game 1	Game 2	Game 3	Game 4	Game 5	Game 6	Mean
1	2.09 (B)	1.89 (E)	1.92 (A)	2.00 (C)	1.82 (F)	2.15 (D)	1.98
2	2.00 (D)	3.00 (A)	4.00 (E)	4.00 (B)	4.00 (C)	4.00 (F)	3.50
3	2.50 (A)	4.00 (F)	2.60 (D)	1.86 (E)	1.82 (B)	1.82 (C)	2.43
4	2.60 (F)	3.00 (C)	2.56 (B)	2.00 (D)	3.50 (A)	1.82 (E)	2.58
5	3.00 (E)	2.40 (B)	2.17 (C)	3.00 (F)	2.40 (D)	1.88 (A)	2.47
6	4.00 (C)	3.25 (D)	3.20 (F)	4.00 (A)	2.50 (E)	4.00 (B)	3.49
Treatments	A: QOI High, Shared viz		C: QOI Med, Shared viz		E: QOI Low, Shared viz		
	B: QOI High, Post viz		D: QOI Med, Post viz		F: QOI Low, Post viz		

Team	Game 1	Game 2	Game 3	Game 4	Game 5	Game 6	Mean
1	3.25 (B)	3.40 (E)	4.00 (A)	4.00 (C)	3.00 (F)	4.00 (D)	3.61
2	3.40 (D)	3.50 (A)	3.50 (E)	2.40 (B)	4.00(C)	2.67 (F)	3.24
3	3.12 (A)	4.00 (F)	4.00 (D)	2.67 (E)	3.40 (B)	3.50 (C)	3.45
4	4.00 (F)	4.00 (C)	4.00 (B)	3.12 (D)	3.75 (A)	3.75 (E)	3.77
5	3.25 (E)	2.17 (B)	3.25 (C)	4.00 (F)	4.00 (D)	3.50 (A)	3.36
6	32.60 (C)	4.00 (D)	4.00 (F)	4.00 (A)	3.86 (E)	3.850 (B)	3.66
Treatments	A: QOI High, Shared viz		C: QOI Med, Shared viz		E: QOI Low, Shared viz		
	B: QOI High, Post viz		D: QOI Med, Post viz		F: QOI Low, Post viz		

SSA scores, agent-based experiment

ANOVA of SSA scores for human-based experiment

	Sum of squares	Degrees of freedom	Mean square	F statistic	p-value
Team (row)	11.48	5	2.30	5.25	0.003
Game (column)	0.37	5	0.07	0.17	0.97
Treatment	2.90	5	0.58	1.33	0.29
QOI	0.84	2	0.42	0.96	0.40
Med/Low - High	0.84	1	0.84	1.91	0.18
Med - Low	0.0001	1	0.0001	0.0002	0.99
Visualization	0.46	1	0.46	1.06	0.32
Interaction	1.60	2	0.80	1.83	0.19
Error	8.75	20	0.44		
Total	23.50	35			

	Sum of squares	Degrees of freedom	Mean square	F statistic	p-value
Team (row)	1.18	5	0.24	0.83	0.54
Game (column)	1.10	5	0.22	0.77	0.58
Treatment	1.73	5	0.35	1.22	0.34
QOI	0.55	2	0.27	0.97	0.40
Med/Low - High	0.12	1	0.02	0.41	0.53
Med - Low	0.43	1	0.001	1.53	0.23
Visualization	0.06	1	0.06	0.22	0.64
Interaction	1.12	2	0.56	1.98	0.16
Error	5.67	20	0.28		
Total	9.68	35			

ANOVA of SSA scores for agent-based experiment

Accuracy scores, human-based experiment

Team	Game 1	Game 2	Game 3	Game 4	Game 5	Game 6	Mean
1	0.35 (B)	0.41 (E)	0.44 (A)	0.67 (C)	0.5 (F)	0.43 (D)	0.47
2	0.67 (D)	0.92 (A)	0.67 (E)	1.00 (B)	1.00 (C)	0.67 (F)	0.82
3	0.15 (A)	1.00 (F)	0.62 (D)	0.23 (E)	0.35 (B)	0.60 (C)	0.49
4	0.62 (F)	0.80 (C)	0.39 (B)	0.67 (D)	0.86 (A)	0.45 (E)	0.63
5	0.33 (E)	0.67(B)	0.77 (C)	0.83 (F)	0.75 (D)	0.27 (A)	0.60
6	1.00 (C)	0.62 (D)	0.50 (F)	1.00 (A)	0.50 (E)	1.00 (B)	0.77
Treatments	A: QOI High, Shared viz		C: QOI Med, Shared viz		E: QOI Low, Shared viz		
	B: QOI High, Post viz		D: QOI Med, Post viz		F: QOI Low, Post viz		

Team	Game 1	Game 2	Game 3	Game 4	Game 5	Game 6	Mean
1	0.38 (B)	0.41 (E)	0.33 (A)	0.67 (C)	0.33 (F)	0.67 (D)	0.47
2	0.65 (D)	0.38 (A)	0.38 (E)	0.75 (B)	0.33 (C)	0.06(F)	0.42
3	0.32 (A)	1.00 (F)	0.40 (D)	0.00 (E)	0.31 (B)	0.43 (C)	0.41
4	1.00 (F)	0.50 (C)	0.67 (B)	0.32 (D)	0.73 (A)	0.47 (E)	0.61
5	0.69 (E)	0.38 (B)	0.31(C)	0.25 (F)	0.67 (D)	0.38 (A)	0.45
6	0.62 (C)	0.67 (D)	0.00 (F)	0.67 (A)	0.30 (E)	0.14 (B)	0.40
Treatments	A: QOI High, Shared viz		C: QOI Med, Shared viz		E: QOI Low, Shared viz		
	B: QOI High, Post viz		D: QOI Med, Post viz		F: QOI Low, Post viz		

Accuracy scores, agent-based experiment

ANOVA of accuracy scores for human-based experiment

	Sum of squares	Degrees of freedom	Mean square	F statistic	p-value
Team (row)	0.62	5	0.12	4.58	0.006
Game (column)	0.26	5	0.05	1.93	0.14
Treatment	0.69	5	0.14	5.17	0.003
QOI	0.54	2	0.27	10.12	0.0009
Med/Low - High	0.38	1	0.38	14.09	0.001
Med - Low	0.16	1	0.16	6.16	0.02
Visualization	0.0002	1	0.0002	0.009	0.92
Interaction	0.15	2	0.075	2.80	0.08
Error	0.54	20	0.03		
Total	2.10	35			

	Sum of squares	Degrees of freedom	Mean square	F statistic	p-value
Team (row)	0.19	5	0.04	0.73	0.61
Game (column)	0.33	5	0.07	1.027	0.31
Treatment	0.47	5	0.09	1.81	0.16
QOI	0.44	2	0.22	4.20	0.03
Med/Low - High	0.20	1	0.20	3.89	0.06
Med - Low	0.24	1	0.24	4.50	0.05
Visualization	0.003	1	0.003	0.06	0.80
Interaction	0.03	2	0.02	0.30	0.74
Error	1.04	20	0.05		
Total	2.04	35			

ANOVA of accuracy scores for agent-based experiment

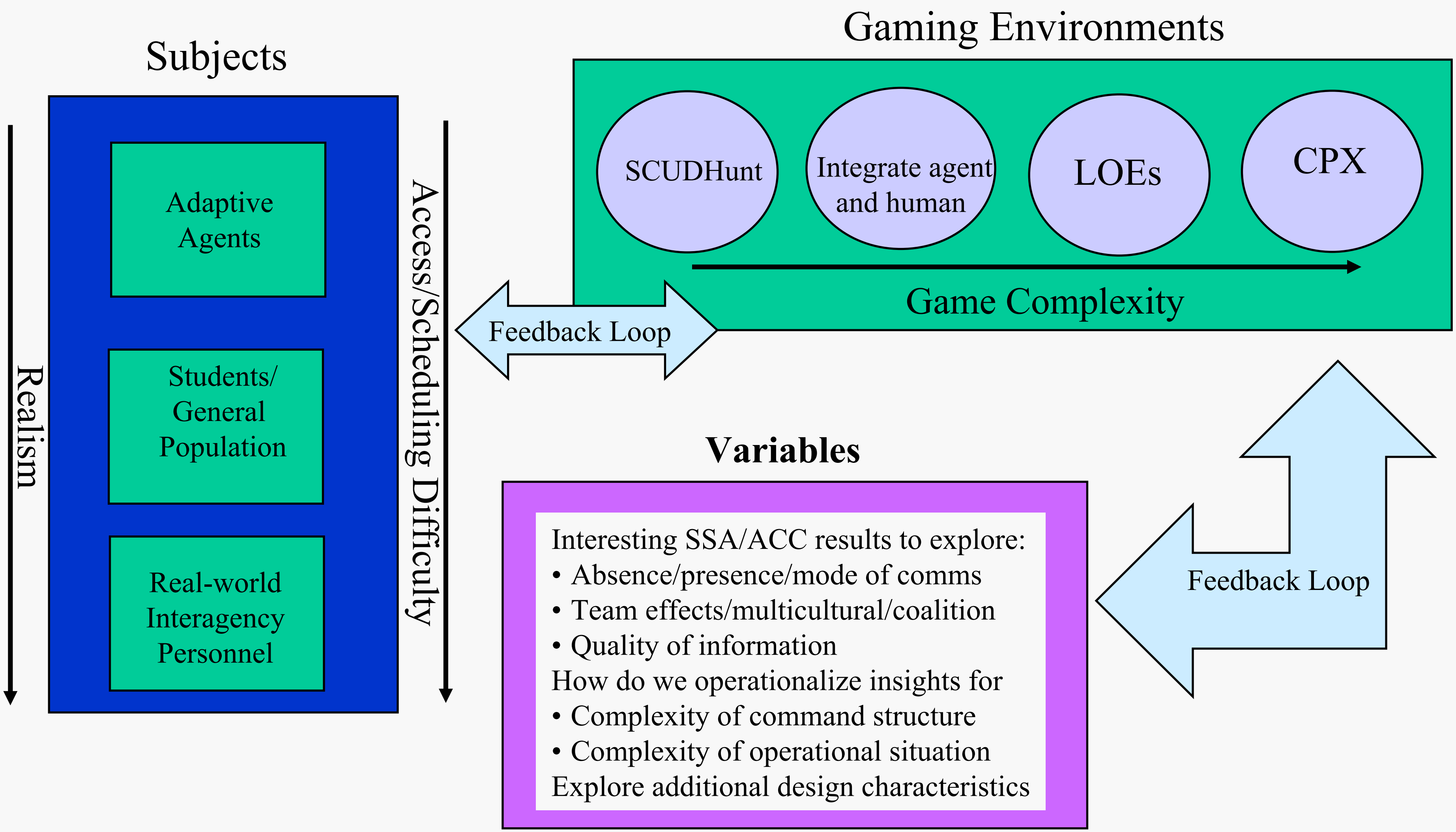
Questions for further research

- The causality conundrum: does high SSA lead to high quality, or does high quality produce high SSA?
- How does adding complexity change the problem (thinking OPFOR, terrain cues)?
- What information do teammates exchange to produce effective SSA and good decisions?
- What attributes of players and teams relate to higher quality scores?
- What is the role of leadership in building SSA and improving quality of decisions?

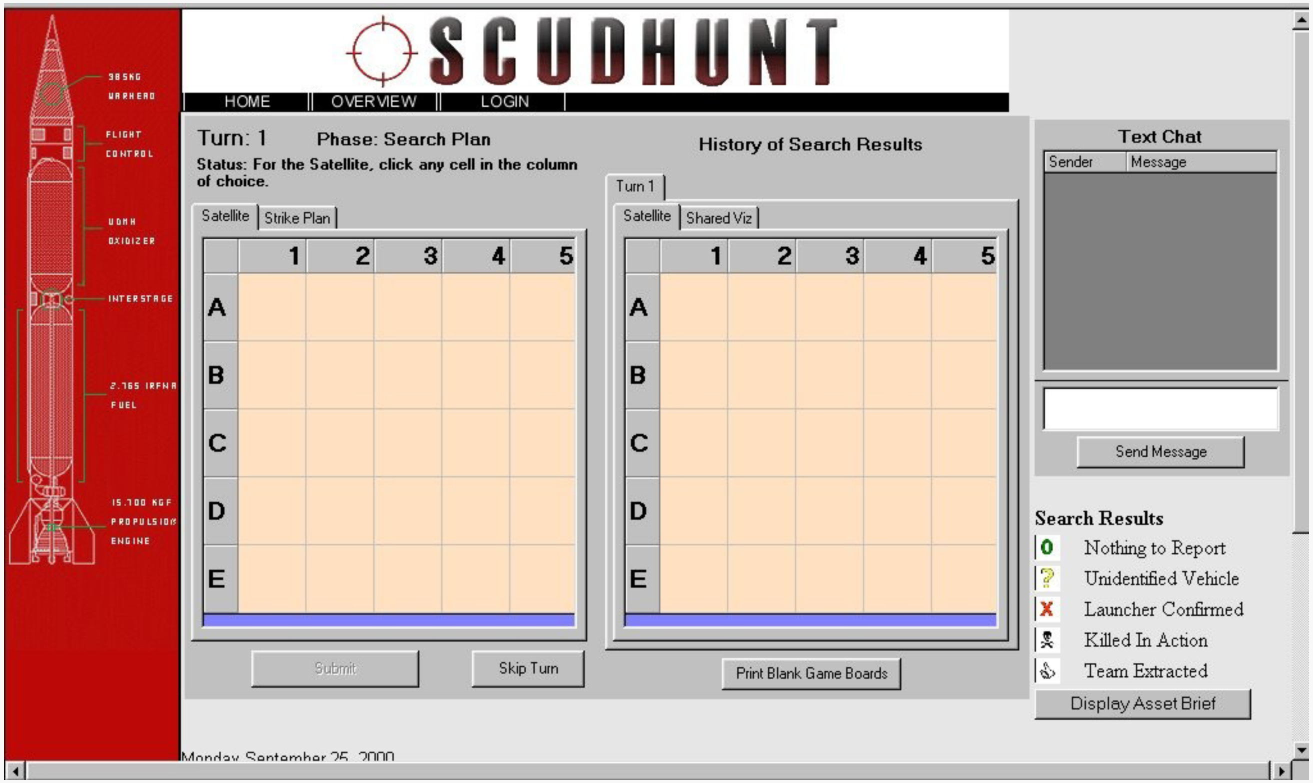
Summary . . . so far!

- We can create distillation games that capture the key elements in the OODA loop
- We can use such games to create experiments that are amenable to statistical design and analysis
- We can use game-playing agents and genetic algorithms to explore vast C2 decision spaces
- We can use human games to validate findings, suggest adjustments, and identify new areas for exploration
- We can integrate agent and human games in experimental campaigns to address fundamental issues systematically

C2 campaign plan



To play SCUDHunt for yourself, go to:
www.scudhunt.com



To read SCUDHunt papers go to:
www.thoughtlink.com/publications.htm

Agent basics

- State of the game
  - *Belief-matrix*, $-1 \leq B_{ij} \leq +1$
- Agent characteristics (~ “Personality”)
  - *Interpretation of sensor reports*
  - *Trust (of other agents)*
  - *Strike Plan Logic*
  - *Sensor Placement Logic*

Agents and sensors

- Interpretation of sensor reports
  - *Sensor-Report:Launcher-Correlation Matrix:*

    β_{RS} = *Agent's belief that launcher is at coordinate for which sensor S has reported R*

  - *Sensor Reliability Estimate Matrix:*

    $\mathcal{R}_{RS}$ = *A's estimate of the reliability of sensor S's report R*

    $0 \leq \mathcal{R}_{RS} \leq 1$

Sensor placement

- Sensor Placement Logic
  - Dogma Threshold, $0 \le B_{Dogma} \le 1$:
    - ✓ *If $B_{ij} \ge B_{Dogma}$ then A places a "launcher is definitely here" marker at site (i,j)*
    - ✓ *If $B_{ij} \le - B_{Dogma}$ then A places a "launcher is definitely not here" marker at site (i,j)*
  - Sensor Placement Fitness Function:

$$F_S(t) = \begin{cases} w_{MCov} * (\text{number of sites covered at time } t) \\ \quad + w_{CCov} * (\text{total number of sites covered at least once for times } t < t) \\ \quad + w_{FCov} * (\text{minimal number of sites that can be covered at time } t+1) \\ \quad + w_{GBel} * (\text{belief gain throughout battlefield at time } t) \\ \quad + w_{LBel} * (\text{belief at site } i, j \text{ at time } t) \end{cases}$$

Trust and beliefs

- Trust (of other agents)
  - *Agent↔Agent Trust Matrix:*

  $0 \leq T_{AB} \leq 1$

  T_{AB} *= 0: agent A mistrusts everything agent B tells it*

  T_{AB} *= 1: agent A believes everything agent B tells it*
- Belief Update:
  - Own Sensors: $B_{own} = \mathscr{R}_{RS} \bullet \beta_{RS}$
  - Linked Sensors: $B_{linked} = T_{AB} \bullet \mathscr{R}_{RS} \bullet \beta_{RS}$ *or* $B_{linked} = T_{AB} \bullet B_{L,ij}$,

    where $B_{L,ij}$ *is the belief matrix of agents linked to A*

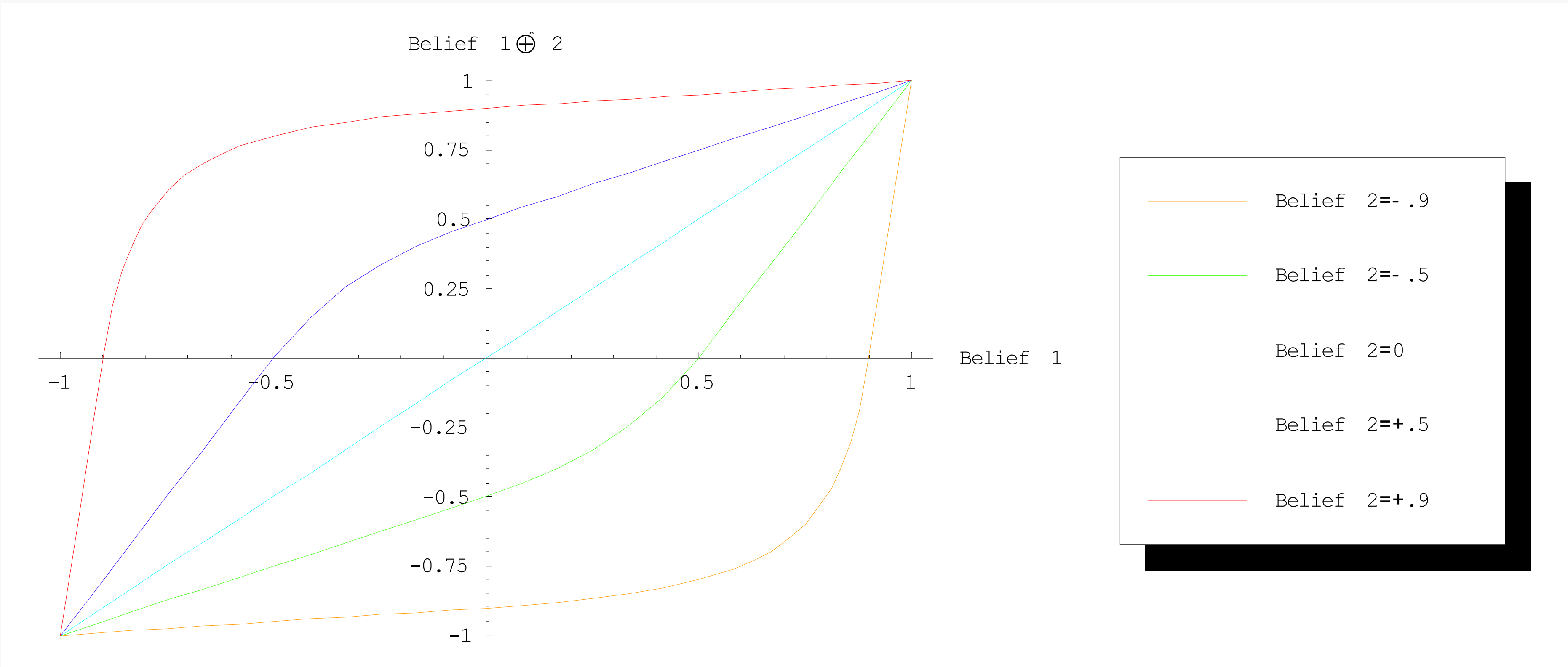
Updating beliefs

- Belief Update (using Durkin fuzzy-sum):

  $B_{ij}(t+1) = B_{ij}(t) \oplus B_{own}(t) \oplus B_{linked}(t)$, where

$$B_1 \oplus B_2 = \begin{cases} B_1 + B_2(1 - B_1), \text{ if } B_1, B_2 > 0, \\ B_1 + B_2(1 + B_1), \text{ if } B_1, B_2 < 0, \\ (B_1 + B_2)/(1 - Minimum\{|B_1|, |B_2|\}) \end{cases}$$

Durkin sums


Belief 1 ⊕̂ 2
1
0.75
0.5
0.25
-0.25
-0.5
-0.75
-1
-1
-0.5
0.5
1
Belief 1
Belief 2=-.9
Belief 2=-.5
Belief 2=0
Belief 2=+.5
Belief 2=+.9

Agents and strike plans

- Strike Plan Logic
  - Select top N_{Strike} ranking sites:

…such that $|B_{ij}| \geq B_{threshold}$

where $0 \leq B_{threshold} \leq 1$ *is A's Threshold Belief Strength*

$B_{threshold} \approx 0 \leftrightarrow$ *A is easily convinced*

$B_{threshold} \approx 1 \leftrightarrow$ *A is stubborn*